DOCUMENT RESUME

ED 207 446 HE 014 395

AUTHOR Breland, Hunter M.
TITLE Assessing Student Characteristics in Admissions to Higher Education. A Review of Procedures. Research Monograph Number 9.
INSTITUTION College Entrance Examination Board, New York, N.Y.
PUB DATE 81
NOTE 143p.
AVAILABLE FROM College Board Publications Orders, Box 2815, Princeton, NJ 08541 ($10.95).

EDRS PRICE MF01 Plus Postage. PC Not Available from EDRS.
DESCRIPTORS *Admission Criteria; Admissions Counseling; *College Admission; *College Applicants; Higher Education; Innovation; Measures (Individuals); *Personality Assessment; *Student Characteristics; *Student College Relationship; Student Interests; Student Motivation

ABSTRACT

Student characteristics such as character, motivation, personality, maturity, creativity and judgment are examined to evaluate the relative utility of various procedures used in college admissions, to suggest possible new ways to use these procedures, and to explore less common procedures. The basic structure of this review is taken from that of Willingham (1974), in which graduate school admissions procedures were analyzed with respect to reliability, validity, and acceptability. Acceptability covers such considerations as implementation, fairness, and public acceptance. Procedures examined include: biographical data, recommendations, interviews, interest measures, personality measures, and combined procedures. The following research activities are offered as suggestions for improving admission procedures: (1) a comprehensive questionnaire; (2) improved reference recommendation procedures; (3) improved interview procedures; (4) formal interest assessment; (5) paramorphic representations of judgments; (6) reliability of admissions procedures; (7) criterion development; and (8) development of new procedures to help solve admissions problems. Appendices include biographical data, recommendations, admissions interviews, interest measures, and personality measures. Tables, references, and an annotated bibliography are provided. (LC)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED207446

Assessing Student Characteristics in Admissions to Higher Education

A Review of Procedures

Research Monograph Number 9

by Hunter M. Breland

Research Psychologist
Educational Testing Service

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☑ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL IN MICROFICHE ONLY HAS BEEN GRANTED BY

CEEB

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

HE 014 395

The College Board, New York, 1981

Editorial inquiries concerning this book should be addressed to: Editorial Office, The College Board, 888 Seventh Avenue, New York, New York 10106.

Copies of this book may be ordered from College Board Publication Orders, Box 2815, Princeton, New Jersey 08541. The price is $10.95.

Cover photograph by Glenn Foss

Copyright © 1981 by College Entrance Examination Board. All rights reserved.

Library of Congress Catalog Number 81-68275

ISBN: 0-87447-138-9

Printed in the United States of America

9 8 7 6 5 4 3 2 1

Contents

Tables iv

Abbreviations v

Preface vii

Introduction 1

Biographical data 5
- Reliability of biodata 5
- Validity of biodata 6
- Acceptability of biodata 10

Recommendations 13
- Reliability of recommendations 14
- Validity of recommendations 14
- Acceptability of recommendations 17

Interviews 18
- Reliability of interviews 19
- Validity of interviews 20
- Acceptability of interviews 23

Interest measures 24
- Reliability of interest measures 24
- Validity of interest measures 25
- Acceptability of interest measures 28

Personality measures 29
- Reliability of personality measures . . . 30
- Validity of personality measures 31
- Acceptability of personality measures . . 34

Other procedures and measures 35
- Creativity measures 35
- Cognitive style measures 38
- Interpersonal competence measures . . 39
- Peer ratings 39
- Parent ratings 42
- Leaderless group discussion 42
- Work samples 42
- Personal statements 43
- Assessment centers 45
- Minicourses 46

Combined procedures 47
- Multiple prediction of academic outcomes 47
- Multiple prediction of nonacademic outcomes 51

Summary and conclusions 53
- Reliability 53
- Validity 53
- Acceptability 54
- Summarization of considerations 56
- Research recommendations 56

References 59

Annotated bibliography 68

Appendixes 97

A. Biographical data 97

B. Recommendations 107

C. Admissions interviews 113

D. Interest measures 119

E. Personality measures 126

Tables

1. Selected admissions procedures and outcomes. 3
2. Predictive validity of biographical inventories 7
3. Incremental validity of biodata 8
4. Examples of biographical inventory items and factors. 9
5. Secondary school achievements for entering college freshmen, fall 1968. . 11
6. High school activities from the National Longitudinal Study, spring 1972. 12
7. Predictive validity of recommendations . 15
8. Inter-rater agreement in interviews . . . 19
9. Group vs. individual interview agreement 20
10. Predictive validity of interviews. 21
11. Predictive validity of interest measures. . 26
12. Reliability of selected personality measures 30
13. Predictive validity of personality measures 31
14. Incremental validity of personality measures 33
15. Predictive validity of creativity measures 37
16. Predictive validity of peer ratings 41
17. Multiple prediction of academic outcomes 48
18. Multiple prediction of nonacademic outcomes 50
19. Predictive validity comparisons of selected procedures. 54
20. A heuristic summary 56

Abbreviations

A . . . . . .academic
AACH. . . . . .Academic Achievement Scale
ACE. . . . .American Council on Education test
ACL . . . . . .Adjective Check List
ACT . . . . .American College Testing Program
AIM. . . . . .Academic Interest Measures
BEQ. .Background and Experience Questionnaire
BI . . . . . .Biographical Inventory
BQ . . . . . .Biographical Questionnaire
CPI . . . . .California Psychological Inventory
CPT . . . . . .California Performance Test
CMHA. . . . .California Mental Health Analysis
CSQ . . . . . .College Student Questionnaire
DAT . . . . . .Dental Aptitude Test
EPPS. . .Edwards Personal Preference Schedule
ERQ. . . . . .Employment Recommendation Questionnaire
F. . . . . .female
FYA. . . . . .first year average
GPA. . . . . .grade point average
GRE . . . . . .Graduate Record Examination
GMAT . .Graduate Management Admissions Test
HS . . . . . .high school
HSA. . . . . .high school average
HSR . . . . . .high school rank (record)
KPR . . . . . .Kuder Preference Record
KR. . . . . .Kuder-Richardson
LGD. . . . . .leaderless group discussion
LSAT . . . . . .Law School Admission Test
LEI . . . . . .Life Experience Inventory
LSF . . . . .Law School Recommendation Form
M . . . . . .male
MBTI . . . . . .Myers-Briggs Type Indicator
MCAT . . . . .Medical College Admissions Test
MG . . . . . .motivation for grades
MMPI . . . .Minnesota Multiphasic Personality Inventory
NA . . . . . .nonacademic
NMSS . . . . . .National Merit Survey Scales
OAIS. . . . .Opinion, Attitudes, Interest Survey
OBI. . . . . .Objective Behavior Inventory
OII. . . . . .Occupational Interest Inventory
OPI. . . . . .Omnibus Personality Inventory
PBI. . . . . .Personal Background Inventory
pbis. . . . .point-biserial correlation coefficient
PF. . . . . .Personality Factors (Cattell)
PH . .Personal History (Worthington) technique
phi . . . . . .phi coefficient
Q by Q . . . . . .question by question
QPA . . . . . .quality point average
rbis . . . . . .biserial correlation coefficient
REF . . . . . .Resident Evaluation Form
SAT-M . . . . . .Scholastic Aptitude Test, Mathematical sections
SAT-V. . . . . .Scholastic Aptitude Test, Verbal sections
SCAT. . . . . .School and College Ability Tests
SCI . . . . . .Selection Council Index
SCII. . . . .Strong-Campbell Interest Inventory
SDS . . . . . .Self-Directed Search
SI. . . . . .Student Interest (rating)

SIAS. . . . .Student Interest and Attitude Study
SLF . . .Standard Law School Recommendation Form
SOV. . . . . .Study of Values
SSHA. . . .Survey of Study Habits and Attitudes
STEP . .Sequential Test of Educational Progress
SVIB . . . . . .Strong Vocational Interest Blank
TAQ . .Talented Accomplishments Questionnaire
TAT . . . . . .Thematic Apperception Test
TDOT . .Thorndike Dimensions of Temperament
TGI. . . . . .Test of General Information
VPI . . . . . .Vocational Preference Inventory
x-valid . . . . . .cross-validation sample

Preface

This review was conducted as one of the initial activities of the Personal Qualities in Admissions Project.[1] This project, sponsored by the College Board and Educational Testing Service, has as one objective the exploration of ways in which colleges might better assess those student characteristics that are not primarily academic in nature but are of importance in the admissions process. Such explorations have been of long-term interest to the sponsors over the past 20 years. Because of the social context in which admissions work is now conducted, however, the Personal Qualities in Admissions Project represents work in an important new direction.

One element that is new is the recognition that institutional objectives extend far beyond maximizing student academic performance. Accordingly, in seeking evidence for the usefulness of an admissions procedure one is not limited to the narrow view that prediction of probable academic performance is all that is important or even that such performance is primarily what is important. There are numerous other objectives of higher education including the creation of stimulating learning environments, the development and nurturance of students, the promotion of nonacademic achievements, the encouragement of student persistence to a degree, and the satisfaction of societal needs for certain kinds of abilities and interests. Thus, in the evaluation of selection procedures, a demonstration that a procedure predicts academic performance is neither necessary nor sufficient as a justification for the use of the procedure.

Despite the new directions that admissions work is taking, in reviewing past research on admission procedures one is limited to the research that is available. Most of that research has attempted to correlate quantitative indexes of student characteristics with quantitative indexes of student performance. Of course, most of these correlations are between traditional academic admissions indexes (high school grades and test scores) and short-term academic outcomes (e.g., freshman grade point average). It is not the objective of this book to examine such relationships, since these have been well researched and reviewed, but to examine other procedures that have been studied in an admissions context and that may offer promise for improving admissions practices. The emphasis is on five procedures for which evidence proved to be most accessible: biographical data collection, recommendations, interviews, interest measures, and personality measures. Several other procedures are also described briefly, and evidence related to admissions is presented.

1. For a description of this project, see "The Case for Personal Qualities in Admissions," by Warren W. Willingham, *The College Board Review*, No. 116, Summer 1980.

Although the primary interest of the sponsors was on college admissions, the available evidence from that one setting was limited. As a result, evidence from an assortment of settings was used: graduate admissions, professional school admissions, and employment. Since the outcomes of interest were not limited to academic performance, evidence from many settings was relevant. Research on medical school admissions proved to be a rich source of data.

Because of the considerable scope and the diverse sources of the research reviewed, no claim is made that this review is exhaustive. Specialists on any one of the admissions settings or on any one of the several procedures reviewed will undoubtedly be able to point to work that has been missed. But such specialists may also find evidence from outside their domain to be intriguing if not enlightening.

A number of people contributed to the collection of information represented in this book. An earlier but brief exploration, under the direction of Thomas L. Hilton, provided a large base of papers and technical reports. W. B. Schrader contributed also to that initial project, and David Reggi provided research assistance. Shula Minsky and Philip A. Griswold located and abstracted additional information after the project became a part of the Personal Qualities project. Francine Mittleman prepared the manuscript with occasional help from Jessie Jones and Debra Smolinski of the Office of Program Research at Educational Testing Service.

Introduction

The term "student characteristics" suggests a broad range of traits, such as character, motivation, personality, maturity, creativity, judgment, and the like—and such traits are indeed of primary importance in this review. In practice these characteristics are most often assessed through a variety of procedures—such as personal history information, letters from those who know the applicant, interviews, and other techniques—rather than as unitary traits. The objective of this review is to evaluate the relative utility of the various procedures that are now used, to suggest possible new ways to use these procedures, and to explore less common procedures.

Previous reviews of a similar nature have had a somewhat different focus. For example, one extensive review by Lavin (1965) was exceptionally broad in scope, covering all levels of education, intellective factors, personality factors, and sociological determinants. Other reviews have been limited to specific admissions settings such as those of Fishman and Pasanella (1960) and Lenning et al. (1974) on college admissions, Willingham (1974) on graduate school admissions, and Cuca et al. (1976) on medical school admissions. Another review technique has been to focus on specific types of procedures as in the reviews of the interview by Ulrich and Trumbo (1965) and Wright (1969), and the review of biographical data of Freeberg (1967). A slight variation on the procedural theme is the review of the self-report by Baird (1976) in which biographical data, interests, attitudes, and any other information self-reported by the applicant were examined. There are difficulties with all of these approaches. Those limited to specific procedures at times lean either toward advocacy or toward criticisms of a procedure. Moreover, consideration of a single procedure in isolation precludes the possibility of noting complementary roles for different procedures or of assessing the relative utility of each in comparison with competing alternatives. The most comprehensive approaches tend to be unmanageable and often result in serious omissions in several areas.

The present review attempts to strike a balance between these extremes. It is probably most akin to the reviews of Fishman and Pasanella (1960) and Willingham (1974)—with the exception that it does not highlight traditional academic variables such as academic records and test scores. However, the limitation in these two reviews to a specific setting (college admissions in the first, and graduate school admissions in the second) excludes rich sources of experience obtained in medical schools, law schools, and other admissions settings. Since the procedures themselves are similar whatever the setting, it seems appropriate that information concerning them should be collected from as many settings as possible. In casting such a wide net there is, of course, the

risk of superficiality. But this problem is muted somewhat by careful attention to and use of the reviews of others who are experts in the various specialties. Since experiences gained in one setting are often not communicated across disciplinary bounds, this approach should assist in communication between admissions specialists in different kinds of institutions.

The basic structure of the review is taken from that of Willingham (1974), in which graduate school admissions procedures were analyzed with respect to reliability, validity, and acceptability. These three features are, of course, essential in any setting. *Reliability* reflects a concern with the consistency with which a measurement or a judgment can be made. At a more practical level, a procedure that is reliable can be repeated for the same individual and yield the same outcome within reasonable limits. If a procedure is not reliable, it is not likely that it can be shown to be valid.

Validity refers to the usefulness of a measure or a judgment. Most commonly, a procedure is considered useful if it predicts some outcome of importance—as college grades are predicted by high school grades, for example. But predictive validity is only one way of demonstrating the utility of a procedure. Other methods are construct validation, in which evidence is assembled to support a procedure as a measure of some definable trait, and content validation, in which a procedure is shown to have content that is of unquestionable importance in a given situation. Validity can also be demonstrated by showing that a procedure is effective in helping an institution achieve certain stated goals. For example, an institution may wish to introduce diversity into its student body, maintain student persistence to a degree, or encourage certain career choices. Or an institution may be primarily concerned with student development, as in the apprenticeship approach to higher education suggested by Wallach (1976).

One of the principal difficulties in predictive validation is the problem of the criterion, the outcome to be predicted. An average grade index for the first year of an academic program, whether it is the freshman year of college or the first year of law school, is a convenient criterion because it is a commonly developed index and because one does not have to wait long (from the time the predictor data were collected) to obtain it. Longer-range criteria are victims of the vicissitudes of time, unanticipated events, staff turnover, and the like. It is not surprising, therefore, that fewer predictive studies are conducted with longer-range criteria. While these longer-range studies are desirable, the faculty in institutions of higher education, as elsewhere, are very much concerned with the performance of students when they enter an institution and the amount of teaching effort and time required to impart the essentials of what is to be taught. The importance of performance during the first year coupled with the relative ease with which such performance data can be obtained probably explain why there have been so many studies of it, but long-range criteria—academic as well as nonacademic—are also important.

Criteria also vary with respect to their reliability as measures. The first-year grade average is a relatively reliable measure in comparison to similar grade averages obtained in later years (e.g., the senior year in college) or in the later stages of professional programs of education (e.g., the clinical years in medical school). If the criteria are not reliable measures, it is difficult to demonstrate statistically that admissions indicators are related to them—and thus are valid as predictors of later performance. Such an inability to demonstrate predictive validity may not mean so much that the admissions indicators are invalid as that the criteria are unreliable. A good example of this situation is the subjective judgments that medical school faculty use as the criterion of performance in the clinical (usually the

last two) years of medical school The subjective judgments tend to be inconsistent from one faculty member to another and are thus difficult to predict. The grade average in the senior year of college, like that in graduate school, also tends to be a poor criterion. The reason is that there is usually little variation in such higher-level grade averages. In summary, the predictive correlation obtainable when studying admissions indexes is quite often as much a function of the quality of the criterion as it is of the quality of the predictor. Accordingly, some caution is necessary in the interpretation of correlation coefficients and other statistics used in evaluating the predictive validity of admissions procedures. Nevertheless, a reasonably good positive correlation suggests reliable measurement on both the predictor and the criterion side. When such positive correlations are repeated over a number of studies they take on important meaning.

The final consideration in this review, *acceptability*, covers a number of practical considerations that govern the use of procedures. These include economic feasibility, practicality of implementation, fairness, and public acceptance. Any one of these practical considerations may tend to dominate decisions concerning usage, at times inappropriately. Extreme economic constraints may force an institution to a narrow set of procedures that may favor some segment of the applicant pool. Conversely, widespread public acceptance of a procedure may exist even though no evidence exists to support the validity of the procedure. More often, however, it is the lack of public acceptance that tends to prevent the use of certain procedures.

A number of procedures, some common and some not so common, are considered in the following sections with respect to their reliability, validity, and acceptability. The order in which the

Table 1. Selected admissions procedures and outcomes

	Criteria (outcomes)									
Procedures	*Academic*	*Non-academic*	*Persistence*	*Career choice*	*Diversity*	*Equity*	*Privacy*	*Public acceptance*	*Enrollment*	*Economy/ workability*
Biographical data										
Recommendations										
Interviews										
Interest measures										
Personality measures										
Other procedures										
Combined procedures										

procedures are considered (see Table 1) reflects roughly an ordering by degree of usage. One begins with biographical data because that is essentially what the application blank collects, and the application blank is the most fundamental procedure of all, in either admissions or employment selection. The application blank can easily be extended to become a biographical inventory with additional questions and scaling procedures. Recommendations are probably the second-most-common procedure in admissions work. Interviews, while common, are not usually required of all applicants (except in employment settings), and their use varies somewhat. Interest measures are taken next because even though they are less common when used as formal measures, information about interests is often obtained through admissions materials. The final procedure listed in Table 1, personality measurement, is rarely used in practice, but much research on the use of such measures in admissions has been conducted. Several other little-used procedures are covered briefly but are not reviewed in any formal sense. These other procedures include creativity measures, peer ratings, parents' judgments, work samples, interpersonal skill measures, assessment centers, and minicourses. Combined procedures are considered last.

To the degree that it is possible, each of these procedures is considered in the context of a broad range of outcomes, as indicated in Table 1. Much past research has concentrated on the prediction of academic performance, but special efforts have been made to uncover research studies of the prediction of nonacademic performance, persistence, and career choice. In recognition of the limited perspective offered by predictive objectives, special efforts have also been made to consider other objectives of higher education: diversity, equity, privacy, public acceptance, enrollment, and economy. All of these are legitimate goals in admissions decisions. As conceptualized in Table 1, any of the procedures that contributes to the attainment of any of the desired outcomes is a useful procedure. However, some procedures may contribute more than others.

Biographical data

Biographical data are used in admissions in a variety of ways. The application blank itself is used to collect biographical information on applicants. The biographical inventory is essentially an extension of the application form, even though it may be used separately and given a different name. The information collected may be weighted in different ways and empirically keyed to predict some outcome of interest. Some biographical inventories contain items on interests, attitudes, and other areas, but the treatment of the more formal interest and attitude inventories will be kept separate in this book. However, since brief questions reflecting interests and attitudes may be included in biographical inventories, they are to some degree represented in this chapter.

There are good reasons why biographical data are frequently collected and used. They are easily obtained because the applicant expects that such information will be collected, and they are inexpensive to collect and process, easily handled by information-processing equipment, and usually accurate. The information can be used in numerous ways beyond aiding in admissions decisions. For example, it can be used to help institutions identify special kinds of applicants and it can be used to inform applicants of special programs and institutions that may be of interest to them. This kind of information exchange may become increasingly important as institutional and societal interests in admissions shift from a dominant concern with future academic performance to broader concerns of nonacademic performance, persistence, career choice, diversity, and student development.

Given the wide range of questions that can be asked in an application blank or a biographical inventory, any evaluation of the reliability, validity, and acceptability of such questions or combinations thereof must be carefully attuned to the different types of information encompassed. Some questions are essentially demographic, but even within this category there are important differences. Questions about gender and ethnicity, for example, are of proven utility and have come to be accepted as appropriate. Other questions are at times more personal, not necessarily useful, and possibly offensive. Parents' income, for example, may contribute to the prediction of academic performance in an empirically keyed inventory, but it does not follow that it is an appropriate item for use in making admissions decisions.

Reliability of biodata

Owens (1976), one of the leading advocates of the use of biodata in selection, observes that "any question regarding the reliability of biodata inventories or scales is a subject for equivocation." The problem with reliability estimation of biodata scales is that they tend to be more heterogeneous

than homogeneous, and, accordingly, estimates of internal consistency may be misleading. Items for scales are often selected because they correlate with an external criterion of interest, not because they correlate with other items. Owens reports test-retest reliabilities for the Alpha Biographical Inventory, developed at the Institute for Behavioral Research in Creativity, of between .82 and .88 for ninth-grade students. Also described by Owens are three other studies. Chaney and Owens (1964) readministered an 82-item inventory to college students 19 months after an initial administration, and a retest estimate of .85 was obtained. Bryson (1969) obtained average retest estimates over a one-week interval of .90 using a sample of college males. Baehr and Williams (1967), however, reported internal consistency estimates from .43 to .76 in a sample of 680 vocationally heterogeneous males.

Nicholson (1970) reports a coefficient of stability of .81 for a personal information form administered to 80 college applicants and then repeated eight months later for those accepted and enrolled at a selective private college. Also reported by Nicholson was a test-retest coefficient of .84 for an accomplishment inventory over the same time span.

Baird (1976), another advocate of biodata assessment, argues that while traditional concepts of reliability apply to scales used in biographical questionnaires, they are generally difficult to apply to individual item measures. He notes that the exception is test-retest reliability, but even here the formulas are not always appropriate because many items do not form scales in the usual sense. To Baird the essential issue of importance is whether respondents make the same response each time an item is presented, and such a view seems the appropriate one for considerations of the reliability of biographical data. Certainly there is no necessity for homogeneity in an application blank. For certain kinds of questions one can expect very high repeat consistency.

Validity of biodata

Fishman and Pasanella (1960) in their review of the decade 1949–1959 noted that the biographical inventory had commanded considerable activity during that period. A total of 23 studies were cited in which biodata were used to predict college grades. Predictive correlations for these studies ranged from .01 to .63, with a median of .13. It was emphasized, however, that few of these studies included predictor variables usually available, such as the high school record and test scores, to determine if substantial gains in prediction could be attributed to biodata. One study by Myers (1952) was cited as evidence that the high school average, when added to a test score, eliminated any independent contribution of biodata. Both the Myers (1952) study and another by Malloy (1955) had suggested that biodata added significantly to the prediction possible with a test score alone. Malloy and Ivanoff (1964), Aiken (1964), and Payne et al. (1973) have since completed studies supporting the contention that biodata add practically and significantly to the prediction of college grades. On the other hand, Webb (1960), Willingham (1965), and Hilton and Myers (1967) came to the opposite conclusion from their analyses.

In his review Freeberg (1967) suggested that biodata would have to be related to broader criteria before their potential could be realized. Some promise for biodata was indicated in a study by Anastasi, Meade, and Schneiders (1960) in which a comprehensive and long-range criterion of college performance correlated .55 in a cross-validation with appropriately keyed biodata. Similar success was reported in studies of National Merit Scholarship applicants (e.g., Nichols, 1966; Nichols and Holland, 1963). High school achievements and activities were some of the best predictors of college achievement in leadership, science, music, and art—even when the group studied was already selected at a high level

in academic variables. Some indication that biodata might predict persistence was noted, even though studies by both Willingham (1965) and Prediger (1965) offered little evidence for such.

Baird (1976) makes a particularly strong case for the use of certain types of biodata in his monograph on self-reports. In addition to reporting much information on predictive relationships between biodata and outcomes of importance, special attention is given to the authenticity question of whether biodata can be trusted. Additionally, biodata are offered as one approach to the issue of fairness in selection for higher education. Owens (1976) and McCormick and Ilgen (1980) report

Table 2. Predictive validity of biographical inventories

Study	Outcome predicted	Predictor	Sample	Predictive correlations[a] A	NA
Aiken (1964)	Freshman GPA, withdrawal	76-item BI	100F	.57	
Anastasi et al. (1960)	Comprehensive college performance	80-item BI	300		.55
Anastasi & Schaefer (1969)	HS art ach.	166-item BI	100F		.34
	HS writing ach.		100F		.55
Harrington (1969)	Freshman GPA	300-item BI, scale 1	746M	.27	
		scale 2		.21	
		scale 1	780F	.30	
		scale 2		.17	
Lunneborg & Lunneborg (1966)	Withdrawal	5-item BI	520		.35
	Change of major	8-item BI			.35
Malloy & Ivanoff (1964)	Sophomore GPA	Life Experience Inventory	60M	.56	
			80F	.52	
Payne et al. (1973)	Freshman GPA	389-item BI (acad. ach. scale)	931M	.45	
			818F	.41	
Schaefer (1969)	HS writing ach.	165-item BI	400M		.64
Scott & Castles (1975)	HS talented accomplishments	118-item BI	123M		.33
			144F		.67
Webb (1960)	Freshman GPA	200-item BI	303M	.41	
			172F	.45	
Willingham (1965)	Freshman GPA	Application blank	1,242M	.48	
	Voluntary withdrawal				.25
Median correlations				.43	.35

a. "A" indicates academic outcomes and "NA" nonacademic outcomes.

Note: See Appendix A for further details of these studies.

much success with the use of biodata in industrial psychology. High correlations are reported predicting overall job performance, managerial success, scientific creativity, and employee turnover. Wing and Wallach (1971) have also demonstrated how students' previous accomplishments predict accomplishments of the same type in college.

Despite the generally positive nature of these reviews, it is of interest to look more closely at the validity issue. Table 2 presents representative data in the form of zero-order correlation coefficients from a number of recent studies, some of which were noted by the reviewers. Both academic and nonacademic outcomes are considered, even though nonacademic outcomes are in some cases inextricable from academic outcomes. All figures reported are based on cross-validation samples, except those of Payne et al. (1973) and Malloy and Ivanoff (1964). The median correlation for academic prediction is lower than the level of prediction possible, on the average, when traditional academic predictors are used to predict short-term academic outcomes. However, the biodata inventories are probably superior to traditional measures as predictors of nonacademic outcomes.

Table 3 summarizes incremental validity analyses reported in five studies. The studies by Payne et al. (1973) and Malloy and Ivanoff (1964) report the largest increments, but these may be due to capitalization on chance, since no cross-validation procedure was described. Only one of these studies, reporting a substantial incremental validity attributable to biodata, used a cross-validation sample. Thus, the evidence on incremental validity of biodata is not convincing.

Table 4 shows predictive correlations for individual items or types of items of biodata inventories from several studies. Only the best predictors have been included in Table 4 as examples. There are two things of particular interest in this table. First, what are termed biodata items are at

Table 3. Incremental validity of biodata

Study	Outcome predicted	Incremental predictor	Baseline predictors	Sample	Predictive correlations R_1	R_2	ΔR
Aiken (1964)	Freshman GPA	76-item BI	HSR, SAT, class size	100F	.58	.70	.12
Payne et al. (1973)	Freshman GPA	389-item BI, acad. ach. scale	HSA, SAT	931M 818F	.40 .44	.56 .53	.16 .09
Malloy & Ivanoff (1964)	Sophomore GPA	Life Experience Inventory	HSR, ACE, Z[a] HSR, ACE, CMHA[b]	60M 80F	.57 .65	.69 .70	.12 .05
Webb (1960)	Freshman GPA	200-item BI	HSA, math test, ACE	303M 172F	.65 .59	.66 .60	.01 .01
Willingham (1965)	Freshman GPA	35-item application blank	HSA, SAT	1,242M	.60	.62	.02

a. Z-inventory score based on a subset of 72 MMPI items.

b. California Mental Health Analysis score.

Note: See Appendix A for further details of these studies.

Table 4. Examples of biographical inventory items and factors

Study	Outcome predicted	Predictor item	Sample	Predictive correlations[a] A	NA
Brown & Dubois (1964)	Freshman GPA	Lowest acceptable GPA	125 engr.	.56	
		Hours study per week	students	.30	
Korman et al. (1968)	Internship success (faculty rating)	Frequency of teen dating	68 med.		.31
		Many friends	graduates		.27
		Class or club officer			.24
		Independence encouraged			.22
		Parental interest			.28
	Peer esteem (peer rating)	Not raised in large city			.34
		Class or club officer			.33
		Parent-child discussions			.25
		Not school behavior prob.			.25
	Humanism (faculty rating)	Not raised in large city			.36
		Independence encouraged			.31
		Talked with parents			.27
		No conflicts with parents			.26
Richards, Holland, & Lutz (1967)	College leadership	HS leadership ach.	2,792 college		.29
	" music ach.	HS music ach.	sophomores		.41
	" artistic ach.	HS art ach.			.41
	" writing ach.	HS writing ach.			.43
	" science ach.	HS science ach.			.31
Nichols & Holland (1963)	College leadership	No. of HS offices held	544M		.34
			275F		.20
	" literary achievement	HS artistic ach.	544M		.14
			275F		.36

a. "A" indicates academic outcomes and "NA" nonacademic outcomes.

Note: See Appendix A for further details of these studies.

times actually surrogates for traditional academic predictors. For example, in the study by Brown and Dubois (1964) self-reported "lowest acceptable GPA" (grade point average) was a good predictor ($r = .56$) of the academic outcome (freshman GPA), but it is likely that self-reported academic achievement correlated highly with actual achievement as represented by the usually available high school record. Second, while some of the items making nonacademic predictions seem quite appropriate, others may not be. For example, would it be appropriate to reject medical school applicants from large cities because they are found to lack humanism or peer esteem on graduation from medical school? These difficulties need not be taken as prohibitive—since inappropriate items can be removed—but they do suggest that care should be taken in the selection of biodata items.

These validity data for biographical inventories and individual items are promising. However, much work appears to be needed in choosing the

most appropriate biodata items to include. Additionally, Brumback (1964) has pointed to the possibility that extraneous factors may operate to contaminate biodata validity coefficients and, consequently, that reported values could be inflated. For example, biodata items may ask for criterion information known or easily predictable by the respondent. Even if predictive validity coefficients are inflated, however, biodata can be useful as a source of data for information systems needed for the study of retention and recruitment and for information exchanges between students and institutions.

Acceptability of biodata

If properly designed and used, there is no question about the economic feasibility of biodata as an adjunct to the regular admissions process. They are easily adapted for computer applications. In fact, these kinds of data are already extensively used for student description. In a sense, it is only a small step to extend the present materials to more sophisticated applications. Much work is needed, however, to determine which specific biodata items are appropriate for selection purposes.

The procedure also seems to be acceptable to the public generally, since applicants routinely complete application forms and lengthier biographical inventories as part of the process of entering higher education in a number of different settings. But inventories of excessive length and those with offensive or inappropriate questions will very likely encounter difficulties.

The most troublesome aspect of self-reported data of any type has been the belief among admissions officials and admissions researchers that such data lack authenticity or that they are fallible. It is contended that—were it known that data of this type were important in making selection decisions—applicants would learn sooner or later what the correct responses are. Admissions counselors could become even more knowledgeable, and some might pass the information on to favored applicants. Baird (1976), in defense of self-reported information, has argued that the authenticity issue is less important than some believe it to be. Studies by Walsh (1967, 1968, 1969), Baird notes, have demonstrated that self-reported information is generally accurate. Walsh (1967) also studied the effects of incentives to distort responses and concluded that they had little effect on accuracy. However, if questions are excessively personal, they tend to yield distorted responses (Walsh, 1968). Other studies have shown that self-reported grades, for example, correlate highly (.96) with actual grades (Nichols and Holland, 1963).

The acceptability of biodata items depends somewhat on what kind of information is at issue. It is well known that socioeconomic status and similar measures correlate with academic performance, but it does not follow that they are appropriate as considerations for admission. Moreover, it is well known that biographical information is used for affirmative action and other preferential purposes (e.g., geographic diversity). A concern about data on accomplishments is that those who have not had opportunities for accomplishment may be penalized. Disadvantaged students, for example, may not have attended schools where there were adequate opportunities to engage in some activities because of lack of facilities or for other reasons. If a particular honor society does not have an active chapter in the school attended, it is not possible to have been initiated into the society. But the evidence available indicates that disadvantaged students do about as well in high school accomplishments as do other students. Baird (1967) compared college-bound student groups in the lowest quarter of parent income to those in the upper 1 or 2 percent of parent income and found very small differences in real-life accomplishments. In the higher levels of achievement, virtually no differences were observed. In a second study (Baird, 1969c), students

from families of different income levels were found not to differ significantly in the number of accomplishments reported or in the level of accomplishments reported. The evidence available so far is indicative of few mean differences for self-reported accomplishments for different ethnic groups.

Bayer and Boruch (1969) sampled black and white students in 358 colleges and found that slightly greater proportions of black students participated in extracurricular activities and had more nonacademic achievements than white students. An excerpt from the Bayer and Boruch study is shown in Table 5, where the proportions of students who responded "yes" to a number of questions about their high school achievements are indicated. While there are some noticeable differences to specific questions (few women received varsity sports letters in 1968, and black males tended not to get elected to a scholastic honor society), the *average* responses at the bottom of the table are quite similar. In fact, black males come out slightly ahead of the other groups despite their tendency not to get elected to scholastic honor societies. Results of a similar type, shown in Table 6, were obtained in the National Longitudinal Study (Hilton and Rhett, 1973), with the exception that black women were not favored by the average responses to an activities inventory.

One charge by Fishman and Pasanella (1960) and others has been that the use of biodata has no theoretical foundation and thus that it is less acceptable to the research community. McCormick and Ilgen (1980) in a recent discussion confirm that biodata have acquired a poor reputation among scientific industrial psychologists. But recent theoretical work by Owens and Schoenfeldt (1979), it is noted, has begun to change this image. Extensive data on 2,000 college students were subjected to factor analysis to identify clusters of students based on profile similarities and factor scores. These clusters were then re-

Table 5. Secondary school achievements for entering college freshmen, fall 1968

Item	Percentage responding "yes"			
	Men		Women	
	Black	Nonblack	Black	Nonblack
Elected president of student organization	26.1	19.8	25.6	20.3
High rating in state music contest	10.1	8.1	10.3	12.3
Participated in a state or regional speech contest	5.6	5.0	7.9	6.3
Had a major part in a play	24.4	15.8	25.7	16.9
Won a varsity sports letter	48.4	45.4	10.8	13.4
Won an award in art competition	7.6	4.4	4.7	6.5
Edited newspaper, yearbook, or magazine	8.6	8.0	14.0	15.4
Had original writing published	10.9	12.7	15.2	20.1
Participated in NSF Summer Program	1.9	1.0	1.4	.6
Placed in science contest	4.4	2.4	5.3	2.0
Elected to scholastic honor society	13.9	19.3	28.0	33.6
Received National Merit recognition	5.9	7.2	7.6	7.6
Average response	*14.0*	*12.4*	*13.0*	*12.9*

Source: Bayer and Boruch (1969)

Table 6. High school activities from the National Longitudinal Study, spring 1972

Item	Percentage responding "yes"			
	Men		Women	
	Black	Nonblack	Black	Nonblack
Athletic activities	37.3	34.6	43.4	0.0
Cheerleader, majorettes, etc.	16.7	13.3	4.5	22.4
Debating, drama, band, chorus	33.2	26.4	21.5	32.4
Hobby club activities	16.8	16.0	20.4	11.9
Honorary clubs (Beta, Honor Society, etc.)	9.3	13.2	9.4	16.3
Newspaper, yearbook, annual, etc.	17.2	15.2	11.3	19.4
Science, history, language clubs, etc.	28.8	21.0	17.1	26.3
Student council, government, political club	18.7	12.9	11.9	14.9
Vocational clubs (farming, homemaking, etc.)	25.3	15.9	11.5	22.1
Average response	*22.6*	*18.7*	*16.8*	*18.4*
Average response (excluding athletics)	*20.8*	*16.7*	*13.4*	*20.7*

Source: Hilton and Rhett (1973)

lated to vocational interests, creativity, and leadership activities. Owens and Schoenfeldt also describe an assessment-classification model that is intended to match biodata profiles with job families. Schoenfeldt (1974) examined the validity of a biodata model for the assessment of individual characteristics with a large sample of college students. In this study subgroups formed on the basis of biodata collected in the freshman year of college were shown to differ on several criterion measures obtained four years later. The subgroups differed, for example, with respect to curricular paths taken during college.

A potentially troublesome characteristic of biographical inventories was described by Pace and Schoenfeldt (1977). Since empirical weighting schemes may weight items that cannot be shown to be relevant to the behavior they purport to predict, such items would not be "job-related" as required by federal guidelines for employment selection. To the degree that such employment guidelines may influence educational practices, item relevance is also important in admissions. Guion (1965) has called the weighted application "raw empiricism in the extreme" and has suggested that it may be impossible to unravel all of its complexities sufficiently to make a test of item relevance. Pace and Schoenfeldt, however, argue that content analysis of biodata items solves the relevance problem.

Recommendations

Recommendations come in a variety of forms including ratings and letters of reference. The letter of reference has been for many years almost a sine qua non of applications for both employment and academic admissions. In recent years, however, recommendations have come under increasingly heavy attacks because of the lack of confidentiality brought about by the Family Education Rights and Privacy Act of 1974 (the so-called Buckley Amendment). Possible sex bias, problems of interpretation, the amount of work required for some referees such as secondary school counselors and college professors, and the lack of uniformity of effort on the part of referees are other concerns that have been expressed. Moreover, in a climate of anxiety about declining population rates in some age brackets, there are those who fear that the requirement for recommendations may deter applicants. Nevertheless, recommendations are still widely used.

In an attempt to circumvent privacy restrictions, many institutions advise referees that applicants may sign waivers of their right to access. But at another extreme, referees are also at times reminded that whatever they write may be seen by the applicant and that, therefore, special care should be taken in the preparation of recommendations. These problems with recommendations have at times annoyed those who work with them, both the writers and readers of recommendations. One dean and teacher at an independent day school, after writing 43 letters on behalf of the school and 86 as a teacher, wrote (Wells, 1978): "It depresses me to think that those efforts may not count for much, and even more to wrestle with the question of what constitutes honest disclosure. The colleges encourage 'candid' statements, yet I have acquired sufficient distrust of admissions committees to wonder what they would do with a fully candid statement" (p. 63).

At the other end of the same kind of communication, the director of admissions at a selective Ivy League college notes the lack of consistency in recommendations and thus the unfairness they create for some students through no fault of their own (Wickenden, 1979): ". . . the entire secondary school report for one applicant was: 'Real fine candidate.' Another teacher prepared the same report for *all* applicants [and] made a xerox copy of the report with blank spaces left for the names of students who might ask for recommendations. . . ."

Such concerns are not limited to the undergraduate level. A law school dean of admissions recently wrote (Thomas, 1978): "Letters of recommendation have become so devoid of meaningful content that rarely can an admissions officer put any faith in them. The pattern reveals an inflation of praise, an overstatement of merit, a conspicuous avoidance of anything negative, and a boiler plate

of sanctimony that is reminiscent of Victorian poetry" (pp. 220–221).

And a participant in medical school admissions offered this vignette (Levitt, 1978): ". . . a single bad year, a poor mark in a required course, or a lukewarm letter of recommendation can occasionally turn the tide against an otherwise qualified applicant. In this last regard, I well recall a thirty-minute discussion at one admissions committee meeting about one recommendation: 'I would expect that George will become a good, if not excellent, clinician' " (p. 249). At issue was whether this was a "good" or a "bad" recommendation.

Reliability of recommendations

While there are few research studies of recommendations generally, there are almost none that address the critical question of reliability. Willingham (1974) notes that reliability is the Achilles heel of recommendations and that efforts to improve it have not been successful. Harmon (1966) conducted extensive studies of recommendation used in National Science Foundation scholarship applications for graduate study. Reliabilities of single references were reported in the low .30s, and multiple ratings were not a great deal more successful. Cuca et al. (1976) indicate that medical school recommendations written by undergraduate professors failed to correlate at all with letters written by classmates.

Some evidence on the reliability of recommendations has been obtained in employment settings. Rim (1976) examined the reliability of letters of recommendation written about candidates for teaching positions in a university. Eight-nine letters, about 23 candidates, were analyzed. These letters were then submitted to five different full professors, who were asked to assign ratings of plus, minus, or zero to reflect their judgment of the value of the recommendation for the candidate. For only 14 of the letters did the five judges all give a rating of plus. Twenty of the letters received both positive and negative ratings. It was observed that there was no systematic relationship between the kind of candidate and the kind of letters written about the candidate. It was concluded that the reliability of letters of recommendation leaves much to be desired.

Validity of recommendations

Table 7 summarizes several studies reporting evidence on the validity of recommendations of various types in various settings. The median correlations of .20 and .18 for academic and nonacademic outcomes, respectively, indicate that recommendations are about as useful for predicting one as the other.

Both Calkins et al. (1980) and Richards and Taylor (1961) studied the relationship between reference ratings and medical school performance. The medical school program studied by Calkins was a six-year BA/MD program in which clinical experience is begun in the first year. Thus it is possible to have a clinical performance rating for the first two years of the six-year program, even though normally clinical work does not begin until the third year of medical school. The third-year medical school average correlation in Table 7 reported by Richards and Taylor may more rightfully belong in the "nonacademic outcome" column. No information was given on incremental validity.

Reilly (1976) investigated the relationship between first-year law school performance and eight recommendation form ratings at one law school. Significant predictive correlations were obtained for five of the ratings, but only one (oral communication) contributed significantly to a multiple correlation with undergraduate GPA and Law School Admission Test (LSAT) score. The highest zero-order predictive correlation was for a rating of native intelligence ($r = .30$), which is reminiscent of the relative effectiveness of the interview in the assessment of intelligence. Test scores, of

course, do a much better job in assessing intellectual attributes—as the LSAT score correlation in this case shows. Whereas the recommendation rating of native intelligence correlated .30 with first-year law school average, the LSAT score correlated .43 with the same criterion. The multiple correlation was increased less than .01 by the addition of the Native Intelligence Recommendation rating to the multiple attainable through the use of undergraduate GPA and LSAT score. Ratings

Table 7. Predictive validity of recommendations

Study	Outcome predicted	Type of recommendation or rating	Sample	Predictive correlations[a] A	NA
Browning (1968)	First-year teaching performance ratings	Last supt. ref.	147 teachers		.08
		Last principal ref.	340 teachers		.19
		Last supervisor ref.	96 teachers		.23
		Other supt. ref.	111 teachers		.22
		Other principal ref.	407 teachers		.07
		Other supervisor ref.	53 teachers		.06
		Head of dept. ref.	150 teachers		.21
		Practice teach. ref.	68 teachers		−.03
		Coop. teacher ref.	184 teachers		.09
		Dean of college ref.	32 teachers		.19
		Other references	633 teachers		.18
Calkins et al. (1980)	2-yr. clinical perf. evaluation	Reference ratings	396 nonmin.		.16
		Reference recommend.			.09
		Reference ratings	51 min.		.34
		Reference recommend.			.37
	2-yr. examination	Reference ratings	396 nonmin.	.10	
		Reference recommend.		.10	
		Reference ratings	51 min.	.11	
		Reference recommend.		.03	
	2-yr. GPA	Reference ratings	396 nonmin.	.24	
		Reference recommend.		.11	
		Reference ratings	51 min.	.29	
		Reference recommend.		.20	
	Status in program after 2 years	Reference ratings	396 nonmin.		.14
		Reference recommend.			.07
		Reference ratings	51 min.		.07
		Reference recommend.			.05
Creager (1965)	Ph.D. attainment	Reference ratings	520M	.12	
			191F	.08	
			663M	.18	
			84F	.13	
			585M	.17	
			176F	.21	
			264M	.20	
			769M	.15	
			161F	.23	

Table 7 continued

Study	Outcome predicted	Type of recommendation or rating	Sample	Predictive correlations[a] A	NA
Creager & Harmon (1966)	On-the-job ratings of scientists several years later	Reference ratings	182		.22
			160		.33
			185		.34
			171		.35
			249		.28
			109		.31
			134		.16
			192		.27
			160		.27
			73		.13
Harmon (1966)	Citation counts	Reference ratings	Not described precisely, but large		.11
	Publications counts				.12
Mehrabian (1969)	Grad. school perf.	Letter of rec. rating	79	.36	
	Professors' eval.			.25	
	Content grades			.25	
	Statistics grades			.38	
Reilly (1971)	First-year law school average	Native intelligence	119 law school students	.30	
		Indep. of thought		.16	
		Oral communication		.23	
		Written communication		.27	
		Industry & motivation		.22	
		Judgment & maturity		.26	
		Leadership ability		.11	
		Personality		.13	
Richards & Taylor (1961)	1st yr. med. sch. avg.	Premedical advisory committee rating	57 medical school students	.25	
	2nd yr. med. sch. avg.			.31	
	3rd yr. med. sch. avg.			.03	
Median correlations				*.20*	*.18*

a. "A" indicates academic outcomes and "NA" nonacademic outcomes.

Note: See Appendix B for further details of these studies.

of independence of thought, industry and motivation, judgment and maturity, leadership, and personality all failed to contribute to prediction beyond what was possible with undergraduate GPA and LSAT scores.

Longer range criteria were used in studies of Ph.D. attainment by Creager (1965), graduate school performance by Mehrabian (1959), and research scientist performance by Harmon (1966). For nine different samples of students in nine fields of graduate study, Creager computed point-biserial correlations between reference

ratings and Ph.D. attainment ranging from .08 to .23 with a median of .17. However, incremental multiple correlations (beyond those attainable with GRE and other test scores, but without GPA) were zero in seven of the nine samples. Using the criterion of time lapse between B.A. and Ph.D., Creager obtained zero-order correlations ranging between .11 and .33 with a median correlation of .23. Multiple correlations were not increased in four of the nine groups, but one group had an increase from .40 to .46. However, none of these multiples used undergraduate GPA.

Recommendations have also been studied in nonacademic settings. Browning (1968) obtained a median correlation of .13 between ratings of teaching ability potential for 2,221 teaching job applicants and on-the-job teaching performance ratings. Mosel and Goheen (1958, 1959) obtained modest relationships between employment recommendation ratings and on-the-job supervisors' ratings for a number of job categories. Magnitudes of correlations varied somewhat, however, as a function of type of raters and type of job.

In a survey described by Cuca et al. (1976) it was of special interest to note that recommendations were perceived (by 534 medical school admissions committee members) as most useful for appraising scholastic achievement when applicants from the same premedical college were compared. They were seen as least valuable for assessing motivation or for judging the suitability of the applicant as a potential physician.

These predictive validity studies tend to paint a rather bleak picture for the utility of recommendations as they are normally used. Moreover, there is always the question of what information was available to the referee when making the recommendation. In the case of professors, it might be assumed that they have some information concerning the student's academic performance. And, at times, they may even know the student's test scores. If this is the case, it may be somewhat surprising that the predictive correlations for recommendations appear to be somewhat lower than those possible using grades and test scores alone. Perhaps Frederiksen's (1954) comment about interview ratings (that they tend to water down the information that is already available) is equally applicable for recommendations.

Acceptability of recommendations

Despite apparent unhappiness with recommendations, some evidence suggests that they are indeed used heavily by selection committees. Calkins et al. (1974) report correlations of .38, .38, and .28 in three different years of medical school admissions between reference ratings and a selection committee index used in selection. And Whitla (1968) reports biserial correlations with selection actions in excess of .60 for several years in succession at Harvard College. Recommendations seem to be acceptable for committee decisions even though they tend not to correlate with later performance. These same recommendations may be less acceptable to those who have to prepare them, even though only testimonial evidence is available.

Possible sex bias in recommendations is beginning to generate commentary from women and others (Farley, 1978; Guillemin et al., 1979; Kryger and Shikiar, 1978; Lunneborg and Lillie, 1973), though little concrete evidence of bias has been presented. Another interesting problem of acceptability was studied by Shaffer et al. (1976). It was hypothesized that letters written without waivers of privacy access had less credibility than letters with the waiver. If this proves to be the case, then a possible unfairness in recommendations—brought on by the Buckley Amendment—may have been uncovered.

Interviews

Interviews are used extensively in all kinds of selection situations, and the amount of writing about them is reflective of their popularity and importance. They are used more in employment than in admissions; consequently most research and writing about interviews relates to their use in employment. Few colleges or graduate schools require interviews for all applicants, but medical schools usually interview all applicants being seriously considered for admission.

In an early but important review of the literature on interviews by Wagner (1949) 106 relevant papers or books were located, but only 25 of those reported quantitative evidence; the remainder contained a hodge-podge of contradicting opinions. Several early studies in employment, military, and educational settings are described with reliability and validity coefficients reported. Wagner concluded that much confusion existed as to what can and cannot be accomplished with an interview, that the interview remained popular as a selection procedure despite its questionable reliability, that interviews should be conducted using a standardized form regardless of their purpose, and that an important unresolved question about the interview was whether or not it provided a more accurate appraisal than statistical combinations of other data available.

The next important review on the interview was by Mayfield (1964). Of the 88 citations, only 27 are from the pre-1949 period covered by Wagner. Mayfield reached these conclusions concerning the utility of the interview.

1. A general suitability rating based on an unstructured interview with no prior information provided has extremely low inter-rater reliability.
2. In an unstructured interview, material is not consistently covered.
3. When interviewers obtain the same information they are likely to interpret or weight it differently.
4. Structured interviews, in general, provide a higher inter-rater reliability than do unstructured interviews. In almost all cases, where a satisfactory reliability for the selection interview was obtained, the interview was of a structured form.
5. Although the reliabilities of interviews may be high in some situations, the validities obtained are usually of a low magnitude.
6. When an interviewer has measures of proven validity available, predictions based on the interview and the measures are no more accurate than those based on the measures alone, and frequently they are less.
7. Of the traits that are judged in interviews, only intelligence or mental ability seems to be estimated consistently.

In another review that appeared shortly after Mayfield's review, Ulrich and Trumbo (1965) started where Wagner left off in 1949. It is of interest that these last two almost concurrent reviews contained only 28 common citations and 106 different citations. The major focus in the

later review, however, was on research studies in which the interview was used for predictions of success in private, governmental, or military settings. Special emphasis was given to the distinction between judgments made from face-to-face interactions, judgments made from ancillary data, and judgments made on combinations of information. This distinction serves to introduce the fundamental issue of whether predictions are made best from actuarial or from clinical modes of data manipulation. In agreement with Wagner's conclusions, Ulrich and Trumbo concluded that there was a need for greater standardization in the interview and that a reduction in the scope of the interview coupled with more emphasis on the use of ancillary data in actuarial modes would be desirable. In most of the studies reviewed it was found that data from the interview were inextricably confounded with biographical or other data. Even when it appeared that the interview was the sole basis for a decision, it seemed possible that the same predictive efficiency could have been attained through some other less expensive procedure. With respect to the contention that the interview is a multipurpose device used for recruiting, public relations, and information dissemination, the authors suggested that their usefulness for these purposes also could be studied through empirical methods.

The latest review of the interview identified for the present purposes was that of Wright (1969), who cited 53 titles in his summary of the research since 1964. While this review tended to emphasize procedural matters, some conclusions on research evidence were made. It was reported that, during the period reviewed, structured interviews continued to exhibit the best inter-rater reliabilities but that there was still little evidence of the efficacy of the interview with regard to cost or optimum length. Wright recommended research on the structured interview. It was also observed that, although validity of the structured interview was not conclusively demonstrated, the goal of making it more economical and useful should be pursued because, regardless of its proven worth, the interview is probably here to stay.

Reliability of interviews

It is useful to excerpt some of the reliability evidence from the reviews cited and other available sources. Table 8 shows coefficients (presumably inter-rater correlations) reported by Wagner (1949). For the 22 traits or characteristics judged in interviews, a median inter-rater correlation of .54 can be computed. These figures give the impression that intelligence, sociability, and submissiveness were judged with somewhat greater consistency than other factors.

Table 8. Inter-rater agreement in interviews

Trait or characteristic	Inter-rater correlations
Ability to present ideas	.42
Alertness	.36
Appearance	.34
Assertiveness	.74
Cheerfulness	.60
Curiosity	.37
Energy	.64
Impulsiveness	.23
Industry	.54
Initiative	.57
Intelligence	.87
Judgment	.31
Manner	.38
Neatness	.23
Quickness	.61
Reliability	.36
Self-confidence	.77
Sociability	.84
Submissiveness	.85
Tact	.26
Tenderness	.68
Overall ability	.49
Median correlation	*.54*

Source: Wagner (1949)

Table 9. Group vs. individual interview agreement

Trait	Inter-rater correlations
Attitude	.31
Appearance	.46
Interest	.51
Intelligence	.25
Leadership	.60
Maturity	.32
Motivation	.16
Persuasiveness	.32
Self-confidence	.45
Self-expression	.51
Sociability	.44
Potential	.42
Overall	.10
Median correlation	*.42*

Source: Moore and Lee (1974)

Table 9, however, taken from Moore and Lee (1974) does not support the suggestion that either intelligence or sociability is judged consistently. In the latter study 34 graduate students in a business school were interviewed by one of six experienced interviewers for a bank, and these interviews were recorded on tape. Twenty-four bank managers then viewed the tapes and made further ratings of the same type. An additional set of ratings was obtained based on pooled judgments of three of the managers following a group discussion of the tapes. The correlations between the interviewers' ratings and the pooled group ratings, shown in Table 9, serve as evidence of agreement among raters. In contrast to the correlations of Table 8, one might expect those of Table 9 to be generally higher, since group ratings were involved. But the median of the inter-rater correlations of the 13 traits is only .42, suggesting that the median of .54 in Table 8 may be high. The highest correlation in the later study is for leadership, and intelligence and sociability are much lower than would be suggested by Table 8.

Validity of interviews

Table 10 presents a summary of recent predictive validity data on the interview with particular relevance to the present review. All of the predictive correlations were obtained in higher education settings—either college, medical, or law school. The medians of these academic and nonacademic correlations suggest that neither academic nor nonacademic outcomes are accurately predicted by interviews. This is the case even though interviewers often had access to ancillary data on academic performance and test scores at the time the interviews were conducted. The median correlation for the prediction of academic outcomes is only .15 and that for nonacademic outcomes only .10.

Some of these studies are worthy of more elaboration. In the Asher (1970a), study, which reported the highest predicted correlations, the interview was highly structured and there were 10 independent raters in one case and 8 in the other. Moreover, the questions asked in the interview were often directly related to the outcome to be predicted (e.g., "Do you think you are good at statistics?"). In most admissions situations the outcome to be predicted is not quite so specific; as a result, this one study by Asher cannot be suggestive of potential for the interview.

The Calkins et al. (1974) data are of special interest because the study is the result of a long-term effort at the University of Missouri (Kansas City) Medical School to introduce new variables into the admissions process. Both physician and nonphysician interview ratings were used to predict two-year cumulative GPA, performance on an end-of-second-year examination, and clinical performance. No particular pattern of predictability was evident.

The data reported by Gough et al. (1963) are also important. They represent 14 successive classes at the University of California School of Medicine (San Francisco) and the prediction of

both clinical and nonclinical performance. In these data, no advantage is evident for the prediction of the clinical (third and fourth) years from interview ratings. There is more evidence that faculty ratings of graduates were predictable to some modest degree. But peer nominations by fellow medical school graduates were very poorly predicted by admissions interviews. This is indicated as well in the studies of Kegel-Flom (1975) and Korman et al. (1968).

Table 10. Predictive validity of interviews

Study	Outcome predicted	Interview type or rating	Sample	Predictive correlations[a] A	NA
Asher (1970b)	Statistics course grade	Structured, 10 raters	20 undergraduates	.44	
		Unstructured, 8 raters		.38	
Calkins et al. (1980)	2-yr. clinical perf.	Physician ratings	396 nonminority		.05
			51 minority		.25
	2-yr. examination		396 nonminority	−.01	
			51 minority	−.03	
	2-yr. GPA		396 nonminority	.12	
			51 minority	.19	
	2-yr. status in program		396 nonminority		.09
			51 minority		.07
Gough & Hall (1975)	Clinical performance	Admission interview ratings	331 medical students		.09
			330 medical students		.16
	Academic performance		331 medical students	.03	
			330 medical students	−.06	
Gough et al. (1963)	1st yr. GPA med. school	Admission committee ratings	14 classes (medians)	.01	
	2nd yr. GPA med. school			.04	
	3rd yr. GPA med. school			.05	
	4th yr. GPA med. school			.05	
	Faculty factor I				.07
	Faculty factor II				.10
	Faculty factor III				.24
	Faculty factor I′				.14
	Faculty factor II′				.19
Gough et al. (1964)	Peer rating factor I	Admission interview ratings	81 medical school graduates		−.12
	Peer rating factor II				−.02
	Peer rating factor III				−.03
Keck et al. (1979)	Resident evaluations	Physician ratings	56 interns		.19
Kegel-Flom (1975)	Intern supervisor ratings	Admission ratings	110 interns		.04
	Intern peer ratings				.12

Table 10 continued

Study	Outcome predicted	Interview type or rating	Sample	Predictive correlations[a] A	NA
Korman et al. (1968)	General ach. factor (for 4 years of medical school)	Poise	68 interns	.30	
	Intern success factor (supervisor ratings)	Social awareness			.26
	Peer esteem factor (based on ratings made at graduation)	Poise			.25
	Humanism factor (faculty ratings)	Acceptability as student			.07
Linn & Winograd (1969)	1st yr. law school GPA	Admission ratings	332 law students	.00	
	2nd yr. law school GPA		177 law students	−.05	
	3rd yr. law school GPA		83 law students	−.02	
	4th yr. law school GPA		81 law students	−.03	
Murden et al. (1978)	Internship letter ratings	Maturity	458 medical students		.22
		Rapport			.22
		Nonacademic achievement			.18
		Motivation			.18
Richards & Taylor (1961)	1st yr. med. school GPA	Personal characteristics	65 medical students	.16	
	2nd yr. med. school GPA			.23	
	3rd yr. med. school GPA			.22	
	1st yr. med. school GPA	Chances of success		.23	
	2nd yr. med. school GPA			.23	
	3rd yr. med. school GPA			.15	
	1st yr. med. school GPA	Recommended decision		.22	
	2nd yr. med. school GPA			.16	
	3rd yr. med. school GPA			.32	
Median correlations				*.15*	*.10*

a. "A" indicates academic outcomes and "NA" nonacademic outcomes.

Note: See Appendix C for further details of these studies.

Few of the research investigations reported have manifested any special concern with the problem of contamination or with "halo" effects. In many of the interview procedures the interviewer has a complete file on the applicant. This file includes test scores, previous academic records, and other information. Therefore, some positive correlation between the interview rating and subsequent academic performance would be expected—and yet the predictive correlations are quite small. As observed previously in the discussion of recommendations, Frederiksen (1954)

noted that in his experience the interview rating tended to "water down" better data available to the interviewer. The resulting predictive correlation of the interview rating was often less than the predictive correlation of information available to the interviewer at the time of the interview. Similar validity observations for the interview as a selection device have been made in Europe (Mitter, 1979).

No evidence was located in this review for the incremental validity of interviews.

Acceptability of interviews

The interview is a very costly procedure both to the applicant and to the institution. In the educational admissions context this economic factor serves as a source of unfairness because all applicants cannot afford to travel to the campus. And many apply from great distances—even from foreign countries.

Nevertheless, the interview appears to be quite acceptable to many institutions and to applicants generally as a procedure. Practically all medical school applicants who survive an initial screening are interviewed prior to acceptance. Moreover, Burgess et al. (1972) and Calkins et al. (1974) report data showing reasonably high correlations between selection decisions and interview ratings. Mendel and Tabb (1977) argue that the interview is a valuable recruiting instrument for dental schools. And, undoubtedly, such a use will become more and more important to undergraduate colleges in the next several years as the population of high school graduates declines.

Even without such a stimulus, however, the selection interview seems to be pretty well accepted. Perhaps Kelly (1957a), after deploring the lack of reliability and validity evidence for the interview, best explained the reason for the interview's acceptability in noting that "all evidence suggests that it gives a great deal of satisfaction to the persons who use it" (p. 78).

Interest measures

Questions about student interests are commonly used in admissions. Often the "biographical inventories" covered previously contain questions about interests. Indeed, even application forms often ask a few questions about interests. In this section the focus is on the more formal interest measures, such as the well-known Strong Vocational Interest Blank (SVIB), the Kuder Preference Record, and other similar inventories which have been seriously considered for formal use in admissions work—but more brief assessments of interests are included.

A central issue in interest measurement is whether a lengthy inventory is actually necessary. In the traditional psychometric view, answers to direct questions about interests tend to be unreliable, superficial, or unrealistic. In the case of younger persons, it is believed that insufficient knowledge about careers and the real world of work precludes accurate answers to direct questions about interests. In fact, distortions of career and work activities obtained from television, books, peers, and even parents can often stimulate interest or disinterest in a misleading way. It is for these reasons that scientific approaches to interest measurement request responses to a wide variety of activities of possible interest and these responses are empirically keyed to specific occupations or areas of interests. The most well known of these empirically keyed interest inventories is the SVIB, which has undergone numerous revisions and extensions. The latest version is now known as the Strong-Campbell Interest Inventory (SCII). There are a number of similar inventories, and some of these have been used in research on admission to higher education.

The other basic approach to interest assessment simply asks the respondent to express his or her interests directly. Dolliver (1969) and Whitney (1969), in reviews of the SVIB and other interest inventories, came to the conclusion that the predictive validity of expressed interests was at least as great as that of lengthy inventories. Katz (1972), moreover, points out that the factor structures of expressed interest ratings are virtually identical to those of inventory scales, and that by early adulthood "interests seem to occupy a rather well integrated and coherent territory in the individual's awareness" (page 1465). Baird (1976) also presents a particularly strong case for expressed interests.

Reliability of interest measures

Carefully developed interest inventories have respectable reliabilities. Anastasi (1976) reports median retest reliabilities over a 30-day period for the SVIB in the high .80s and long-term stability correlations in the .60s and .70s. Clendenen (1972) reports test-retest correlations for the SVIB basic interest scales over short time spans of .89 and over long time intervals of .65 (for adults) and

.56 (for teenagers). Clendenen also notes that the SVIB-W, for women, is somewhat lacking in data—available test-retest groups are small, with a median reliability of .68 for two groups (one with a 4-year time interval and the other with a 15-year interval). A test-retest reliability of .53 is noted by Clendenen for freshmen tested 26 years later.

For the Vocational Preference Inventory (VPI), Nichols (1966) obtained internal consistency estimates of .82 for a scale developed for grade prediction and of .88 for a scale developed for the prediction of extracurricular achievement. Baird (1969b) reports reliabilities of from .57 to .89 for college freshmen and from .50 to .89 for women.

For the Kuder General Interest Scales (KGIS), Lohnes (1972) reports that internal consistency estimates for all scales are above .70. For the Kuder Occupational Interest Survey (KOIS), Walsh (1972) notes that the manual reports median test-retest reliabilities of .90 for 25 twelfth grade and 25 college students over a two-week time period, but he adds that additional reliability studies with larger samples are needed.

Validity of interest measures

In his review Lavin (1965) describes several studies of the validity of interest measures. One early study showed that an interest measure predicted grades in corresponding courses with correlations from .50 to .70 and improved the predictions obtained using ability measures alone. However, in another study reported by Lavin the SVIB correlated only .19 with grade point average. Two studies used the physician key of the SVIB in a study of medical student performance. In one of these medical school studies the physician scale did not correlate with academic performance, and the other study found little relationship after ability was controlled. A study of engineering students found that interest scores were correlated to performance, but a study of high school students showed that interest in science was positively related to science accomplishment after ability was controlled. Lavin concluded that, while research on interest measures was weak, they appeared to be useful in predicting academic performance. But the evidence reported by Lavin on interest measures is not convincing.

Baird (1976) makes a strong case for the validity of directly expressed interests as opposed to elaborate measures of interest. He cites evidence that, even though the reliability of the SVIB exceeds that of expressed interests, the predictive validity of the latter is at least as great as that of the SVIB. Where direct comparisons of the prediction of occupation have been made, he observes, in no case was the SVIB as accurate. Baird also points out that Strong himself obtained correlations of .72 for SVIB and .69 for freshman choice of major field as predictors of occupation 19 years later. This is probably not a statistically significant difference. Other studies comparing expressed interests to the Kuder Preference Record, Holland's VPI, and other formal measures are also described, and Baird concludes that it is, in fact, surprising that haphazardly constructed expressed interests questions stand up so well against carefully constructed interest measures.

Table 11 presents a summary of several studies reporting validity coefficients for various interest measures as predictors of both academic and nonacademic outcomes. Both direct and indirect measures of interests are included. The studies of Norris and Katz (1970) and of Nichols and Holland (1963) used direct ratings of interests, and the remaining studies used inventories. No superiority for the inventories is indicated. Overall, the median correlations at the bottom of Table 11 suggest that interest measures are better for predicting nonacademic outcomes (median r = .26) than for predicting academic outcomes (median r = .12).[1]

1. It should be noted that validity studies for interest inventories are often conducted by comparing occupational or other groups. Thus no correlations are generated, and consequently such studies are not suitable for a summary of this type.

Table 11. Predictive validity of interest measures

Study	Outcome predicted	Predictor measures	Sample description	Predictive correlations[a] A	NA
Ayers & Rohr (1974)	Teacher training GPA	Kuder−1	203F (elem. program)	−.13	
		Kuder−6		−.13	
		Kuder−9		−.06	
		Kuder−1	400 (secondary program)	−.24	
		Kuder−6		−.05	
		Kuder−9		.22	
Baird (1969b)	College art ach.	VPI-artistic	2,295M		.30
			2,834F		.22
	College science ach.	VPI-intellectual	2,295M		.28
			2,834F		.32
	College music ach.	VPI-artistic	2,295M		.30
	College writing ach.	VPI-artistic	2,295M		.35
			2,834F		.27
	College drama ach.	VPI-artistic	2,295M		.25
Campbell (1966)	Freshman GPA	SVIB-ach.	250M	.36	
Harrington (1969)	Freshman GPA	SVIB-ach.	746M	.29	
Keck et al. (1979)	Resident eval.	SDS-realistic[b]	50 resident interns		.13
Kegel-Flom (1975)	Intern supervision eval.	SVIB-physician	110 interns		.25
	Intern peer eval.				−.07
Lindsay & Althouse (1969)	Freshman GPA	SVIB-ach.	226M	.10	
			88F	.25	
		CSQ-motivation for grades	226M	.27	
			88F	.29	
Nichols (1966)	Freshman GPA	VPI-grade scale	201M normative sample	.10	
			218F	.13	
			179M x-valid. sample	.11	
			138F	−.12	
	Freshman extra-curricular activities	VPI-extrac. scale	201M normative sample		.22
			218F		.07
			179M x-valid. sample		.18
			138F		−.06
Nichols & Holland (1963)	Freshman leadership	Preference for social occupations rating	544M National Merit Final.		.21
			275F		.14
	Freshman literary achievement	Preference for artistic occupations rating	544M		.40
			275F		.38
	Freshman music achievement	Preference for artistic occupations rating	544M		.22
			275F		.03

Table 11 continued

Study	Outcome predicted	Predictor measures	Sample description	Predictive correlations[a] A	NA
Nichols & Holland (1963) (continued)	Freshman graphic arts achievement	Preference for artistic occupations rating	544M 275F		.28 .40
	Freshman art achievement	Preference for artistic occupations rating	544M 275F		.29 .20
	Ph. D. aspiration as freshman	Preference for intellectual occupations	544M 275F		.36 .26
Norris & Katz (1970)	Freshman grades Freshman interests	AIM (Academic Interest Inventory)	(median of 20 r's) (median of 22 r's)	.12	.34
	Freshman grades Freshman interests	SI (Student Interest self-report rating)	(median of 20 r's) (median of 22 r's)	.10	.40
Median correlations				*.12*	*.26*

a. "A" indicates academic outcomes and "NA" nonacademic outcomes

b. Best of six interest scale predictors

Note: See Appendix D for further details of these studies.

Some of the studies reported in Table 11 also reported evidence on incremental validity. Where the attempt was to improve on the prediction of an academic outcome, the increments obtained were generally unimpressive. Harrington obtained an increment of .01 in the multiple R when the SVIB achievement scale was added to an ACT battery and high school rank in the prediction of freshman GPA. In two separate samples (men and women), Lindsay and Althouse (1969) also obtained increments of .01 in the multiple R when the SVIB achievement scale was added to SAT-V, SAT-M, and high school average. Better results were obtained when the CSQ-1 (motivation for grades) scale of the College Student Questionnaire was used. Increments of .04 and .03 in the multiple R for men and women were obtained using this scale in addition to SAT-V, SAT-M, and high school average. Norris and Katz (1970) examined 10 different major field areas for both the AIM and the SI ratings and for both men and women. The median increment in the multiple correlation was only .01 for the AIM and .02 for the SI ratings, even though only PSAT-V and PSAT-M were used as the baseline (that is, high school rank or grades were not included in the equation). One major field was intriguing, however. In the prediction of art and architecture grades, the increments in multiple R were .06 and .15 for the AIM and the SI ratings, respectively. The Norris and Katz (1970) study also demonstrated that, where the outcome predicted is *interest* in a major field area, both the AIM and the SI ratings added substantially to the multiple R obtained using PSAT-V and PSAT-M. For 22 increments computed for 11 fields and for men and women in each, the median increment added by the AIM was .16 and that for the SI ratings was .23.

Interest measures or direct indicators of interest would also seem to be useful as predictors of persistence, although no good evidence on this point was found in this review. Strong interests in

a particular subject area or occupation seem very likely related to a tendency to remain in that major field of study and to persist to employment in the occupation.

Clearly, interest measures are useful in monitoring and influencing diversity within institutions. Expressions of major field choices are often used to determine if an applicant fits an enrollment need. In this way institutions can match student enrollment and faculty availability if that is an important objective. Another use of interests could be to achieve a balance among different fundamental types of students—such as Holland's six VPI types—or to move an institution in some desired direction with respect to the enrollment of different types.

Acceptability of interest measures

Directly expressed interests have some obvious advantages over formal interest inventories when it comes to acceptability. They are inexpensive to develop, they can be brief enough to add to a routine student questionnaire, and they in no way arouse suspicion in the respondent. Moreover, they require much less of the student's time to complete. Interests can of course be faked, but, if they are used primarily to improve persistence and institutional diversity, the kind of faking required to enhance admission is not altogether obvious. Few applicants would know what responses are associated with persistence or even that improving persistence is an objective in admission. And, unless applicants know about institutional goals, it is not likely that they would know what type of interests are in current favor at an institution. Obviously, the acceptability and workability of interest measures imply some restrictions on the public release of information about institutional policies.

Personality measures

In conventional psychometric terminology, personality measures are instruments that intend to assess emotional, motivational, interpersonal, and attitudinal characteristics of individuals. Although some writers include interests in the domain of personality (e.g., Anastasi, 1976; Messick, 1979), interests have been treated in this review separately because the procedural implications for the use of interests in admissions are somewhat different from those for personality measures. It is one thing to ask an applicant to an educational program what his or her interests are but another to probe more deeply into subtleties of emotions, motives, interpersonal relationships, and attitudes.

In their early review Fishman and Pasanella (1960) devoted only two paragraphs to personality measures. They briefly mentioned early studies of the Rorschach, MMPI, EPPS, Taylor's Manifest Anxiety Scale, and Gough's Hr[1] scale. Although some of these scales tended to correlate with college performance, none used the high school record in combination with aptitude test scores as a control. In contrast to Fishman and Pasanella, Lavin's (1965) review devoted more than 30 pages to personality measures. Measures of achievement motivation, independence, impulsivity, anxiety, introversion, self-image, and adjustment were all considered in relation to academic performance at all levels of education. Lavin concluded that, while some connections between personality variables and academic performance appeared to be evident, "many of the relationships . . . are tenuous at best, and it is undoubtedly true that the state of knowledge regarding the relation between personality variables and achievement is still so tentative that it cannot be used confidently for practical purposes, such as for college admissions" (p. 101).

Three years later Stein (1963) reviewed the literature during the decade 1950–60 that pertained to the relation between personality factors and college success. Noting the extreme diversity in criteria, procedures, and populations in the literature and the consequent difficulty in making comparisons, Stein concluded that the task of filling in the critical gaps in knowledge was monumental. He recommended, as a result of this diversity and complexity, that a research clearinghouse should be established. However, very little information on the prediction of college performance using personality measures was presented.

Messick (1967) also considered personality measurement with respect to its potential as a predictor of college performance. Two major questions were posed: (1) Are personality tests

1. Hr stands for "honor-point ratio" and was intended to predict such.

any good as measures of the purported personality characteristics? (2) What should these tests be used for? The first question was viewed as a scientific one and the second as an ethical one. The scientific question was approached from the perspectives of reliability and validity. The ethical question was approached by posing specific questions concerning the use of personality tests in college admissions. The scientific question was answered by describing evidence on validity and reliability. It was concluded that many measures were available, but none measured up adequately against the psychometric standards applied. The ethical question was not answered directly, but it was suggested in conclusion that, even if empirical validity could be established for a personality measure, this would not be enough. One example of the ethical dilemma was illustrated with a question of this type: If it were found that docile, submissive students received higher grades in college, should a preference be given to such students in admission? Interestingly, a review of the research on personality and academic attainment by Entwistle (1972) did, in fact, conclude that success at the university level was associated with introversion. Entwistle noted, however, that the relationship differs with respect to intellectual level, type of institution, and subject being studied.

In a more recent review Messick (1979) considered the potential educational impact of noncognitive personal characteristics with particular reference to problems in their measurement. Several functions of noncognitive measures were considered including guidance, selection, placement, instruction, and education. The paper emphasizes the positive functions of those measures, but it also recognizes pitfalls that may result in their misuse.

Reliability of personality measures

Given the wide range and assortment of personality measures, it is not surprising that reliabilities vary a great deal from one instrument to the next and, within instruments, from one scale to the next. Table 12 presents both internal consistency and test-retest reliabilities reported for three personality measures that have been used in research on admissions. The measures most commonly used in admissions research appear to have reasonable reliabilities in the sense of internal consistency; however, test-retest (stability) evidence is spotty.

Table 12. Reliability of selected personality measures

Measure	Internal consistency	Test-retest[c]
Adjective Check List	.80 (KR-21)[a] .79 (KR-21)[b]	.54 (mean) .80 (median, one-week interval)
California Psychological Inventory	.83 (KR-21)[a] .86 (KR-21)[b]	.65 (high school males, one-year interval) .68 (high school females, one-year interval)
Edwards Personal Preference Schedule	.78 (median)	.83 (median, one-week interval) .73 (median, three-week interval)

a. For special scales developed by Nichols (1966) to predict college grades.

b. For special scales developed by Nichols (1966) to predict college extracurricular achievement.

c. From Buros (1965), Buros (1972).

Validity of personality measures

Table 13 presents a summarization of results of 14 published studies in which a number of different samples were used to correlate a personality measure with a later outcome, academic or nonacademic. Some points of explanation in the construction of this table are important. First, the figures shown represent either correlations obtained using all scales of a measure or, where all

Table 13. Predictive validity of personality measures

Study	Outcome predicted	Predictor	Sample description	Predictive r's[a] A	NA
Ayers & Rohr (1974)	Teacher training GPA	SSHA	203F (elem. program)	−.55	
			400 (second. program)	.12	
Brown & Dubois (1964)	Freshman GPA	MMPI – Mf	76 science & hum. students	.28	
		MMPI – Ae	125 engr. students	.24	
Donnan (1968)	Freshman GPA	OAIS – achiever	375 61–62 freshmen	.41	
		personality	1,747 64–65 freshmen	.36	
	Sophomore		375 61–62 freshmen	.41	
	1st yr. survival		375 61–62 freshmen		.23
			1,747 64–65 freshmen		.13
	2nd yr. survival		375 61–62 freshmen		.25
Dunham (1973)	Freshman GPA	EPPS	161 college freshmen	.11	
		Ego ach. index		.17	
		TAT		.22	
Gough & Hall (1975)	Clinical perform. factor	ACL-composite	330 medical students		.13
			331 medical students		.11
			661 medical students		.25
			330 medical students	.06	
			331 medical students	.08	
Himmelstein (1965)	Freshman GPA	MMPI	193 college freshmen	.39	
Keck et al. (1979)	Resident evaluation	OPI – autonomy	56 med. sch. grads.		.24
Kegel-Flom (1975)	Intern sup. eval.	CPI – tolerance	110 interns		.08
	Intern peer eval.				.35
Korman et al. (1968)	Internship success	CPI – ach. conform.	68 interns		.11
	Peer esteem				−.15
	Humanism				−.02
	Internship success	EPPS – nurturance			.01
	Peer esteem	– achievement			.21
		– nurturance			.08
	General achievement	CPI – cap. status		.17	
	Scientist potential	CPI – ach. indep.		.33	

Table 13 continued

Study	Outcome predicted	Predictor	Sample description	Predictive r's[a] A	NA
McCausland & Stewart (1974)	Freshman GPA	SSHA	154 college freshmen	.40	
Nichols (1966)	Freshman GPA	ACL—grade scale	201M Normative sample	.17	
			218F	.18	
		CPI—grade scale	201M	.34	
			218F	.24	
		ACL—grade scale	179M x-valid. sample	.06	
			138F	.22	
		CPI—grade scale	179M	.36	
			138F	.28	
	Freshman extracurricular activities	ACL—extrac.	201M Normative sample		.18
			218F		.03
		CPI—extrac.	201M		.23
			218F		.10
		ACL—extrac.	179M x-valid. sample		.16
			138F		.07
		CPI—extrac.	179M		.22
			138F		.13
Petry & Craft (1976)	Cumulative GPA	TDOT	52 high risk students	.52	
		SOV		.44	
	Prof. course GPA	TDOT		.28	
		SOV		.39	
Stroup (1970)	Freshman GPA	CPI	968M	.43	
			970F	.35	
Watson (1967)	Freshman GPA	CPI	100 college freshmen	.38	
Median correlations				*.28*	*.13*

a. "A" indicates academic outcomes and "NA" nonacademic outcomes.

Note: See Appendix F for further details of these studies.

scales were not used, the *best* correlation reported for any single scale. In other words, large numbers of very small and negative correlations were discarded because in the actual use of any given measure the best scales would be used. As in previous tables, in making judgments about whether an outcome was academic or nonacademic, the bias was to classify it as nonacademic if there was any justification for so doing. The overall median correlations at the bottom of Table 13 suggest that personality measures may be more effective in the prediction of academic outcomes than in the prediction of nonacademic outcomes. The overall median predictive correlation for academic outcomes is .28 and that for nonacademic outcomes .13. The CPI was used

to generate nine of the correlations with nonacademic outcomes. The medians based on this measure alone are .34 (academic) and .12 (nonacademic), and these also suggest that academic outcomes are better predicted by personality measures than nonacademic outcomes.

The direct comparisons in the medical school study by Korman et al. (1968) also reveal the same pattern. In this study, the outcome measures were all factors based on long-term evaluations conducted at graduation from medical school. The nonacademic outcome factors were based on faculty and peer ratings (each graduate rated all other graduates). The predictive correlations for the more academic outcomes are .17 and .33 (general achievement and science and potential), but the median predictive correlation for nonacademic outcomes is only .06. Undoubtedly, the reliability of the outcome ratings is a factor in these kinds of results and, indeed, may be the most generally important influence on the predictability of nonacademic outcomes.

Some of these same studies have also considered the issue of incremental validity, and

Table 14. Incremental validity of personality measures

Study and setting	Baseline predictors	Personality measures	Outcome predicted	Sample	Incremental R's R_1[a]	R_2[b]	ΔR
Donnan (1968) —college	SAT-V, SAT-M, HSR	OAIS	Freshman GPA	1	.56	.63	.07
			Sophomore cum. GPA		.55	.62	.07
			1st-year survival		.29	.34	.05
			2nd-year survival		.33	.38	.05
			Freshman GPA	2	.55	.59	.04
			1st-year survival		.21	.22	.01
Gough & Hall (1975)	MCAT, Premedical GPA, & Interview	ACL	Clinical performance		.46	.47	.01
McCausland & Stewart (1974) — college	ACT composite, HSA	SSHA	Freshman GPA		.78	.81	.02
Petry & Craft (1976) High-risk colleges	SCAT-V, SCAT-M	TDOT	Cumultaive GPA		.49	.67	.18
			Professional course GPA		.49	.57	.08
			Cumulative GPA		.49	.60	.11
			Professional course GPA		.49	.59	.10
Richards & Casey (1975) (1975) — college, marginal students	SAT, HSA	TEAM	Freshman GPA		.56	.58	.03
Stroup (1970) —college	SAT-V, SAT-M	CPI	Freshman GPA		.52	.60	.08
			Freshman GPA		.50	.54	.04
Median increment							.06

a. Based on baseline predictors.

b. Based on baseline predictors plus personality measure.

those amenable to comparison are summarized in Table 14. Although the overall median increment in the multiple R is .06, most of the large increments come from two studies, Petry and Craft (1976) and Stroup (1970). Fishman and Pasanella (1960) emphasized that the high school record should be used in addition to aptitude test scores in the baseline prediction, but Petry and Craft and Stroup examined the increment in R beyond aptitude test scores only. If those two studies are excluded from the median computation a figure of .05 is obtained. More important perhaps is the need for cross-validation in such studies. The studies reported by Donnan, Petry and Craft, and Stroup give no indication that results were cross-validated. Without these studies, the median incremental validity is an unimpressive .02.

In the interests of brevity, the many criticisms of the validity of personality measures from concurrent, convergent/divergent, and construct validity perspectives are omitted here. Good reviews of these validity problems are contained in Buros (1965, 1972) and in Messick (1967).

Acceptability of personality measures

The issue of the acceptability of personality measures for use in admissions was alluded to in the introduction to this section. There are certainly some ethical questions concerning their use. Another important question relates to public acceptance. Few psychological developments have aroused the public's ire more than have personality measures. With an expected decline in the population of college-aged students. some colleges may be reluctant to add an additional hurdle to college going. The use of personality measures could introduce an additional deterrent to applications because of problems of public acceptance. On the other hand, institutions with an excess of applicants may want to consider personality measures as a means of achieving diversity in incoming classes. But even if the public acceptance problems can be overcome, these institutions will still need to be especially concerned about the ethics of using personality measures.

Other procedures and measures

The previous sections have treated several procedures either often used or often examined in research on admissions. There are of course numerous other procedures that have from time to time been proposed for use in admissions or that appear to offer some promise for use in admissions, but this review did not uncover a sufficient number of studies (in which the procedures were used in admissions settings) to allow for generalizable summarizations. Nonetheless, it is important that some of these other procedures be described briefly because they may be the focus of future research. No attempt has been made, however, to review these procedures in any formal sense. Moreover, no attempt has been made to organize the topics other than by commonly referred-to procedures and specific measures. There is overlap between procedures and measures, and some might logically be subsumed under others.

Creativity measures

Although there exists an extensive literature on creativity and how to assess it, it has not been often examined as a predictor of achievement in higher education.

The long-term work of Torrance on creativity, represented in the Torrance Tests of Creative Thinking (TTCT), is one important line of work. Reviewers in Buros (1972) point to a number of problems with the TTCT, however—the most important of which is that there seems to be little validity evidence and practically no predictive validity evidence to support the use of it. On the other hand, reliability is described as "adequate": test-retest estimates are reported in the range .50 to .93 over one- to two-week periods and from .35 to .73 over three-year periods. However, it is noted that the TTCT seems to be related to both academic intelligence and educational achievement. This relationship leads to a question about the degree to which the TTCT is saturated with general academic aptitude. The second major line of work on creativity has been that of Guilford and his colleagues on divergent thinking. This research on divergent thinking is a part of a much larger effort encompassing all intellectual functioning in the structure-of-intellect model. A total of 14 tests of divergent production was developed; these tests are generally applicable at the high school level and beyond. As for the Torrance tests, reported reliabilities are lower than might be desirable, with split-half estimates ranging from the .60s to the .80s. While factorial validity data are available for these tests of divergent thinking, Anastasi (1976) reports that evidence on criterion-related validity is weak. A few reports that have investigated the use of measures of creativity in higher education or have provided evidence relating to their use in higher education are described below, even though the

ability to generalize beyond these reports may be questioned.

The study of Nichols and Holland (1963) examined creativity as a predictor of college achievement in the sample of National Merit finalists included in the summarizations of biographical data, interest measures, personal statements, and personality measures. (Some of those summarizations will overlap with the discussion of creativity because those modes of assessment overlap with the assessment of creativity.) Nichols and Holland used all four of these modes to measure creativity in the following specific measures.

1. Self-reported artistic achievement in high school. This was an 11-item scale on artistic, musical, literary, and dramatic achievements in high school.

2. A breadth of creativity scale based on a 32-item checklist of creative activities ranging from making Christmas cards and combining new words to initiating a club and planning an independent experiment. The number of activities reported was used as the score.

3. A preconscious activity scale based on the notion of preconscious activity as a part of creative performance. This was a 38-item scale with a KR-21 reliability estimate of .70.

4. Preference for artistic occupations. This was a student rating of interest in a group of six artistic occupations and included previously in the section on interest measures.

5. Enjoyment of creative thinking. Students wrote a free-response statement in response to this question: "What single hour in class during your last school year did you like most? Why?" Responses were coded dichotomously as creative thinking or not.

6. Creative activities enjoyed. This was a count of the number of creative activities checked from a 50-item list of high school activities enjoyed.

7. Originality ratings were made by the students themselves, by their mothers, and by their fathers, to obtain a total of three ratings.

Table 15 is a summarization of the results obtained when the creativity measures were correlated with academic and nonacademic performance in college. It is clear from Table 15 that the creativity measures did not predict academic performance in the first year of college for these National Merit finalists but that most did significantly predict the nonacademic composite of leadership, literary, music, dramatic, and graphic arts achievements. To dispel the suspicion that academic performance was probably not predictable in this group (by any measure) because the students were very select and probably went largely to competitive colleges, it is important to note that the correlation between high school rank and college grades was .24 for men and .21 for women (both significant at the .05 level). And to dispel the suspicion that academic predictors probably predicted the nonacademic performance composite as well as the creativity measures, it is important to note that high school rank correlated −.19 and .04 with nonacademic performance for men and women, respectively.

Wallach and Wing (1969) posited that a person's resourcefulness in generating ideas was an important element of creativity. In a study of entering college freshmen, they showed that ideational productivity as well as the uniqueness of the ideas (as measured by a brief test) was associated with nonacademic accomplishments in high school (as recalled retrospectively). Notwithstanding possible design problems in research of this type (for example, those subjects who tended to respond with many ideas on a test of ideational fluency might merely be the same types of subjects who respond with many high school activities—i.e., effusive respondents vs. noneffusive respondents), there is an indication that ideational fluency is related to nonacademic accomplishment. And such would logically seem

Table 15. Predictive validity of creativity measures

Creativity measures	Sample[a]	Outcome predicted[b]	
		Academic	Nonacademic
Artistic achievement in HS	1	ns	.29
	2	ns	.22
Breadth of creative activity in HS	1	ns	.32
	2	ns	.36
Originality of comment on an issue	1	ns	.20
	2	ns	-.05
Preconscious activity scale	1	ns	.24
	2	ns	.17
Preference for artistic occupations	1	ns	.29
	2	ns	.20
Enjoyment of creative thinking	1	ns	ns
	2	ns	ns
Number of creative activities enjoyed, HS	1	ns	.20
	2	ns	.15
Originality: self-rating	1	ns	.19
	2	ns	.16
Originality: mother's rating	1	ns	.11
	2	ns	.21
Originality: father's rating	1	.05	.14
	2	.15	.18

a. In this table the first sample is always men and the second women.

b. The academic outcome is freshman grades. Nonacademic outcomes are a composite of freshman college achievements of these types: leadership, literary, music, dramatic, scientific, and graphic arts. The abbreviation "ns" means nonsignificant.

Source: Nichols and Holland (1963)

to be the case. No evidence is presented, however, on ideational fluency as a predictor of later academic or nonacademic performance.

More recent research on creativity measurement has been reported by Frederiksen and Ward (1978) and Ward, Frederiksen, and Carlson (1980). In this research, an indication of creative or divergent thinking is obtained by asking the subject to formulate hypotheses about a research finding. A graph, table, or some other mode of presentation is used as the stimulus, and the subject may then state as many hypotheses as can be thought of to account for the result presented. Indications are that this kind of assessment is useful for selection to research training programs in a variety of fields. Frederiksen and Ward (1978) reported on the development of a set of Tests of Scientific Thinking (TST) for possible use as criterion measures in research on creativity. Both the quality and quantity of ideas produced are scored. In the study reported the TST tests were administered to 3,500 applicants for graduate study in psychology, and scores in the TST were correlated with scores on the Graduate Record Examination (GRE). Addi-

tionally, follow-up questionnaires were sent to students asking about accomplishments and other matters. It was concluded that the TST tests have reliable variance that is not being measured by conventional tests like the GRE, even though reliability estimates for the TST were low. Relationships with the questionnaire information were generally low, but they were often statistically significant.

Ward, Frederiksen, and Carlson (1980) examined the construct validity of a Formulating Hypotheses (FH) test when both free-response (FR) and machine-scorable (MS) forms were used. Subjects were 174 paid volunteers from 11 different colleges and universities. Scores on the FH, FR, and MS were correlated with both cognitive and personality factors. It was concluded that the MS version would seem to be of no particular value as a supplement to present aptitude tests, except that it might add face validity. However, the FR version was seen as worthy of further consideration.

Cognitive style measures

As summarized by Messick, cognitive styles are "information-processing consistencies reflective of underlying personality trends" (pp. 286–87). There has been a vast amount of research in this general area, and a number of different cognitive styles have been identified. The most widely examined cognitive style has as one pole field dependence and as the other pole field independence. The concept of field dependence/independence was initially formulated to explain observed differences in performance on a task that involved perception of the upright. In this task the subject is asked to judge whether a line is vertical to the ground. It has been demonstrated that there are essentially two modes of responding to such a task: one relies on cues from the environment (field dependence) and the other relies only on internal cues (field independence). It has been hypothesized that performance on tests of field dependence/independence is associated with a socially sensitive, interpersonal orientation as contrasted with an impersonal orientation. A good review of the educational implications of field dependence/independence has been conducted by Witkin et al. (1977a).

In one college-level study of field dependence/independence Witkin et al. (1977b) followed an entire class of over 1,500 students from entry through graduation and application to graduate school. The Group Embedded Figures Test (GEFT) was used to assess field dependence. GEFT scores showed little relation to either high school grade point averages, college grade point averages, or dropping out at the college or graduate application stage. But relatively field-independent students tended to enter majors in science, while relatively field-dependent students tended to enter education. Moreover, students whose college-entry choices were incongruent with their cognitive styles tended to shift to other majors more compatible with their cognitive styles by graduation.

Goodenough et al. (1979) examined cognitive styles in the development of medical careers. Two longitudinal studies were reported: the first was a study of college undergraduates who had expressed an early interest in medicine, and the second was a study of medical students. In the first study the same data base used by Witkin et al. (1977b) was used. The mean GEFT scores of college freshmen who declared themselves to be premedical students were compared with mean GEFT scores of those who did not. No significant differences were observed. However, in the second study, significantly higher mean GEFT scores were found for students of the same group who subsequently applied to medical school. The results suggested that knowledge about a student's cognitive style would be of value to counselors in dealing with students who profess an early interest in medicine.

Interpersonal competence measures

Theory and research in the general area of interpersonal competence dates back to the 1920s and early considerations of social intelligence. Argyris (1962) and others have written much about interpersonal competence, but the most systematic program of research on this topic was conducted by Guilford and his coworkers. Guilford's (1967) theoretical "structure of intellect" concept dealt with distinct kinds of ability concerning behavioral content. These encompassed coping, judging, remembering, and understanding related to such skills as understanding the behavior of others. Two major studies have been completed by the Guilford group, and both are factor analyses of paper-and-pencil tests using pictorial stimuli (O'Sullivan and Guilford, 1975; Hendricks et al., 1969). The first study identified six factors associated with understanding the behavior of others. The second study isolated six factors associated with coping with the behavior of others.

Stricker (1980) notes in a recent review that a wide range of methods for assessing interpersonal competence have been attempted—including self-report inventories, situational tests, role playing, and interviews—but that performance tends to be related to general ability and that the methods tend to be expensive and not amenable to large-scale use. New possibilities through the use of videotape technology, however, are suggested by recent work of Stricker in the development of an Interpersonal Competence Instrument (ICI). The ICI uses the supervisor-subordinate context and places the examinee in the role of the supervisor. An actor takes the role of the subordinate, talking directly into the camera. The examinee watches the screen and responds to brief remarks made by the actor. Stricker reports preliminary studies of inter-rater reliability in the range .80–.90 for total scores, and internal consistency estimates in the range .74–.82 for individual components of the ICI. Correlations with other variables showed that the ICI total scores were related to age, education, and training in psychology, and another measure of social perception; however, they were also appreciably related to scores on a verbal ability test.

Samph and Best (1979) have reported a similar experimental videotape technique being examined for use in the evaluation of medical student interpersonal skills in relating to patients. A 90-minute cassette is used in this technique to assess medical students' observational skills as well as their ability to interpret statements or actions, analyze and synthesize information, and evaluate the behaviors observed. Moderate reliability (.62) is reported for this technique, but a better reliability (.74) was reported by Best and Samph (1979) for a paper-and-pencil technique for assessing medical student interpersonal skills.

In a recent study by Livingston (1980) six tests of nonverbal communication skills were examined in relation to later ratings of interpersonal skills by faculty in graduate programs in psychology and counseling at two different institutions. No consistent correlations between the tests and faculty ratings occurred across the two institutions. One of the tests, however, a Photo Classification Test, correlated consistently with both academic work and an overall effectiveness rating. But the Photo Classification Test did not correlate at all with traditional aptitude test scores. The Photo Classification Test used was a simulation of a role-playing situation.

Peer ratings

One approach to personality assessment is the peer rating or peer nomination; however, much more than personality is usually being assessed with this procedure. It is usually used with groups in training programs toward the end of the program, when trainees are asked to rate all other trainees on certain attributes believed to be important. Alternatively, trainees may be asked simply to "nominate" three or more classmates as

most likely to be successful in the work, more liked by others, or some other attribute. The technique has been extensively used in military training centers such as Officer Candidate School (OCS). Lewin and Zwany (1976), in a recent review of peer ratings, report a number of studies of reliability. Military trainee split-half estimates of about .90 for leadership nominations have been obtained after only a few days of interaction among the trainees. Tupes (1957), in a study of 790 OCS graduates who had known each other for between 3 weeks and 12 months, obtained internal consistency estimates for ratings between .8 and .9 covering the differing degrees of acquaintance among peers. Hollander (1957, 1965) obtained split-half reliabilities in a naval training setting based on ratings obtained on the fifth day after arrival. These reliabilities were estimated for four separate administrations and were all around .95. Waters and Waters (1970) obtained peer ratings from 53 salesmen who had known each other for only three weeks and estimated inter-rater reliabilities between .8 and .9. Therefore, the reliability estimates reported in studies of peer ratings or nominations are usually quite high. However, Smith (1967) demonstrated that a relatively large number of raters is required to obtain these high reliabilities. Smith computed split-half estimates of peer ratings obtained using Cattell's 42 bipolar traits for 61 different samples in which the number of raters ranged from about 6 to over 30. With 15 or more raters, the median split-half estimate was .83. For more than 20 raters, reliability estimates were about the same, but for less than 10 raters split-half estimates ranged from .40 to .70.

A summary table of peer rating/nomination studies adapted from Lewin and Zwany (1976), with modifications and additions, is shown in Table 16. Most of these studies have been conducted in military or business settings, but they seem relevant to higher education admissions because the criteria represent real-life outcomes—and some of these outcomes have been determined over a relatively long term. The median correlation of the predictions of nonacademic outcomes using peer ratings is an impressive .33. The two studies using academic outcomes, Freeberg (1969) and Smith (1967), do not represent a good comparison because the median may be inflated. This is because the study by Freeberg used a mathematics test as a criterion measure immediately after the peers first cooperated in mathematical problem solving for 50 minutes. It seems rather likely that the peer ratings obtained after this session would correlate highly with a test on the same subject. A similar point can be made about the Smith (1967) results. In this study the peer ratings were obtained immediately prior to midterm examinations and after the peers had been together for nine weeks. The criterion was first-year college GPA. Because these students had observed each other during the first year of college, one might expect a higher predictive correlation than if they had not.

As indicated in the discussion of the reliability of peer ratings, a substantial number of peers is required to obtain reliable ratings—probably in excess of 15. In training programs no problem is usually encountered in obtaining these numbers. But a question is raised about the workability of peer nominations in the normal academic admissions situation. It is easy enough to obtain three recommendations, and it might also be easy to obtain three ratings from peers. But, ignoring the problem of whether the peer information obtained in this way can be trusted to be valid, the number of ratings is far from sufficient to yield high reliability.

In a study by Medland, Yates, and Downey (1974) it was suggested that another problem with peer ratings can be the age of the group. In a senior group of officers who were asked their opinion of peer ratings there was high acceptability. But in a junior group of officers there was much more concern that the technique was a

Table 16. Predictive validity of peer ratings

Study	Type of peer rating or nomination	Outcome predicted [a] A	Outcome predicted [a] NA
Amir et al. (1970) Israeli Army	Basic training for noncommissioned officers (NCO)		.78 (officer success)
	NCO course		.79 (officer success)
Downey (1973) Army War College	High-level officers (2 or more ratings)		.29 (senior selection)
Freeberg (1969)	College students (math ability ratings)	.60	
Haggerty (1953)	Military academic ratings		.30 (overall ratings)
Hollander (1965)	Naval officers OCS ratings		.40 (overall ratings)
La Gaipa (1960)	Naval officers: shore		.26
	Naval officers: fleet		.24
Mayfield (1972)	Insurance sales trainee		.29 (1-yr. survival)
			.30 (productivity)
Ricciuti (1955)	Naval officers (Sample 1)		.32
	(Sample 2)		.22
	(Sample 3)		.33
Roadman (1964)	Middle managers		.41 (promoted twice)
Smith (1967)	College students	.45	
Tupes (1957)	Air Force OCS ratings		.52 (officer success)
Waters & Waters (1970)	Sales trainee ratings		.32 (sales success)
Weitz (1958)	Sales trainee ratings		.40 (sales success)
Williams & Leavitt (1947)	Marine officers (Sample 1)		.47 (combat ratings)
	(Sample 2)		.43 (combat ratings)
Wollack & Guttman (1960)	Naval officers		.36 (overall ratings)
Median correlations		*.52*	*.33*

a. "A" indicates academic outcomes and "NA" nonacademic outcomes.

Source: Lewin and Zwany (1976).

popularity contest, and there was apprehensiveness about evaluating friends. Such concerns would probably reach their maximum among adolescents of high school age. Thus, peer ratings would not seem promising for use in college admissions. On the other hand, there is no evidence that they would not be workable if the number of ratings obtained were sufficient.

Parent ratings

Although rarely studied as a means of obtaining information on applications to institutions of higher education, an obvious source of data is parents or other relatives who have known the applicants most of their lives. The creativity measures discussed above (Nichols and Holland, 1963) contained ratings by parents on a student's originality. Other parent ratings also obtained in the same study included mother's and father's ratings of the student's scholarship, perseverance, drive to achieve, dependability, neatness, popularity, athletic ability, self-confidence, self-understanding, leadership, and independence. In this one study, parent ratings generally did not relate as well to college grades and the nonacademic composite as the creativity measures did. Moreover, there was often an inconsistency between the mother's rating and the father's rating of the same attribute. For example, mothers' ratings of dependability for men correlated .19 with first-year college student, but fathers' ratings of dependability of men did not correlate significantly with grades.

Leaderless group discussion

Both Dunnette (1976) and McCormick and Ilgen (1980) mention the leaderless group discussion (LGD) as one of several assessment center techniques, but the LGD has also been used in other contexts. The LGD is a situational exercise in which an assessee may assume a role (e.g., a business consultant to whom data are provided on an organization). The assessee, along with several other "consultants," may then be given the task of helping the company decide the best course of action. Other tasks may include impromptu talks, written statements, or mock interviewing.

Knapp and Sharon (1975) describe LGDs of two types: those with assigned roles and those with nonassigned roles (McCormick and Ilgen characterize these two types as cooperative and competitive). In the nonassigned LGD, participants are given descriptions of problems (usually organizational problems) of some type. As consultants, the participants resolve the problem and submit a written recommendation. In the LGD with assigned roles, each assessee is assigned a role such as vice president of public relations, personnel, manufacturing, etc.

Assessment measures for the LGD can be obtained by trained staff who serve and rate the assessees on several dimensions, by rating written reports, or through peer ratings by the assessees themselves. In one study by Banta (1967) the individual interview was compared with the LGD and found to be equal to the LGD in both reliability and as a predictor of performance. The interviews consisted of seven questions asked during a seven-minute interview. The LGD consisted of six applicants brought together to discuss the same seven questions during a 40-minute session.

In another paper Harris (1964) described procedures for selecting students for admission to a university in Iran. Initially, one-third of the selections were made on the basis of examination scores. A second screening was then conducted using the secondary school record, and selection from the remaining group (about one-third of the initial pool) was done with the LGD. One group was processed each hour in a single room in which observers rated the applicants engaged in discussion.

Work samples

Typing proficiency tests, foreign language proficiency tests, and drivers' license road tests are

all examples of work samples. The work sample is used to reproduce an important element or elements of the actual performance being assessed. Where possible, the same equipment required to demonstrate performance is used. The brief impromptu essay that is used in the assessment of writing skill is an example of an academic work sample.

Karren (1980) has conducted a recent review of work sample techniques as an outgrowth of federal concerns related to selection of federal white-collar entry-level selection. Karren reports that the work sample has been highly effective in predicting both job proficiency and training success. It has been shown to have validities equivalent to a number of alternatives (Asher and Sciarrino, 1974). One study is described by Karren in which a work sample was determined to have higher predictive validity coefficients than two paper-and-pencil tests (Mount, Muchinsky, and Hanser, 1977). In another study by Cascio and Phillips (1979), the use of work samples and simulations was viewed as responsible for reduced rates of turnover, and selection rates for Spanish, black, and white applicants were .57, .60, and .64, respectively. Schmidt et al. (1977) found work samples to be less difficult and more fair for both black and white samples than traditional tests. Moreover, it is reported that the use of work samples has been well received by applicants.

Nonetheless, the work sample has a number of problems. First, the development and administration of work sample tests are expensive. Where large-scale assessment is conducted, such as is most often the case in admission to higher education, work samples may be especially difficult to implement. Another problem with work samples is that they measure only specific skills needed for specific jobs. Accordingly, one is not able to generalize from work samples to broader skills that may be learned later by capable applicants subsequent to selection. It may be that higher education requires broader skills more than it does specific skills.

Personal statements

A relatively common admissions procedure is to ask the applicant to write a brief statement or to enclose something already written with the application materials. These written statements can, of course, come in a variety of forms—essays, autobiographies, and other free-response writing. The written statement as used in admissions often requires that the applicant write about interests, goals, or whatever it is believed will assist the admissions staff in the evaluation of the application. At times the response is an essay on a topic of interest to the applicant. Rarely is the usual statement a complete autobiography, but often the applicant provides important life history information not available in other parts of the application. Free response statements have also been obtained in research studies related to admissions by means of specific stimuli. For example, a subject may be asked to reflect about his or her future career and then to write something. Responses can then be coded in a number of ways, such as whether they provide evidence of a creative orientation. Similarly, subjects may be asked to describe their most enjoyable activity during the past year, and these responses can be coded. At a much more elaborate level, written statements are used in psychological studies as a projective technique in attempts to gauge fundamental psychological attributes (see e.g., Atkinson, 1958; McClelland, 1961). These elaborate procedures have not been considered practical for admissions work, however.

Since written statements as ordinarily used in admissions are rarely scored in any numerical or categorical sense, very little is known about their reliability. As a baseline of evidence, however, some indication of the reliability of judgments of written statements comes from experience in the assessment of writing skill using brief impromptu

essays. Coffman (1966), for example, obtained writing samples from 646 secondary school students and then had these samples scored by English teachers. Essays on five different topics were written, and each of these was scored independently by two different teachers. Estimates were then made of reliabilities for any combination (up to five) of numbers of topics and numbers of readers. Inter-rater reliabilities ranged from .56 (for two raters) to .76 (for five raters) when a single essay was judged. However, estimates of score reliability (reflecting also errors introduced by sampling behavior on only one topic) ranged from .38 (two raters) to .52 (five raters). Maximum score reliability, of course, was obtained with five raters and five topics (.84). This same study, Coffman (1966), is reported in more detail in Godshalk, Swineford, and Coffman (1966). Studies of writing skill, therefore, suggest that when this one attribute is judged, reliabilities are typically quite low unless several readers and several topics are used. Since usually only a single statement or essay is obtained, and since this sample is usually judged by not more than one or two persons, one would expect that reliabilities are usually quite low. When less-well-defined attributes (e.g., maturity, motivation) are being judged, reliabilities are probably even lower.

Higher inter-rater reliabilities have been reported for lengthier scored autobiographies. Annis (1967), in a review of the scored autobiography, reports inter-judge reliabilities of .93 between two analysts reading a 10-page autobiography and .96 for a topical autobiography assessing attitudes. A common method of analyzing the autobiography has been the establishment of relevant categories and the tallying of content with respect to these categories. Entire autobiographies may also be rated on some single dimension, such as level of insight demonstrated, but reliabilities for such global judgments tend to be lower.

High reliabilities have also been reported for TAT-type stories scored in analyses for *n* achievement and other content. Feld and Smith (1958) also report an experiment in which novice scorers were trained to score with high inter-judge reliability. Reliabilities for scoring stories for *n* achievement ranged from .66 to .95 over four scorers and six stories. These reliabilities were in about the same range as obtained by experienced scorers trained over a long period in close association with experts.

There is much less available evidence on the validity of personal statements. Moderate success was achieved in the prediction of medical school performance in a study by Holmes and Hertel (1967). Randomly selected sketches by 69 students at the University of Michigan Medical School were compared with standing at the end of the first year. Predictions were made, at better than a chance level, of upper, middle, or lower class standing. But no evidence is presented of incremental validity beyond that possible with the usual predictors.

Nevis (1954) used a personal history technique to predict the performance of Air Force officers. Ratings were made of personal histories provided by 100 Air Force captains and these were correlated with superiors' ratings made on 14 different scales. The correlations ranged from zero to .52; the author concluded that further use of the personal history technique was warranted.

A rather different approach, mentioned in a previous section on creativity, was used by Nichols and Holland (1963) to elicit evidence of creativity from National Merit Scholarship finalists. Since these students were already carefully selected on the basis of standard academic indicators, the objective was to further discriminate with respect to nonacademic factors. Several techniques were used. A written statement on social problems was scored both for length and for originality. A free-response comment on most-liked high school class hour was coded for creative thinking versus non-creative thinking. These scorings of free-response statements were found

to correlate positively and significantly with nonacademic achievements during the first year of college for some groups. For example, the length of written statements on social problems correlated .35 with literary achievement for men. Creative thinking correlated .34 with literary achievement for men, but −.14 for women. Conversely, creative thinking in free responses correlated .32 with Ph.D. aspirations for women.

Some exploratory research has been conducted recently with what is called the "future autobiography." These are statements such as those indicating what a person expects to be doing 5 or 10 years hence. For example, Tullar and Barrett (1976) asked 36 salesmen to write future autobiographies about goals, interests, concerns, and other future-oriented topics. These autobiographies were then rated with respect to three constructs. One of these constructs was called "agency" and reflected the extent to which subjects saw themselves as the prime agent in determining their future lives. This one construct, agency, correlated positively with sales success with an average of .31 over 10 different sales performance areas.

Written statements are easy to obtain and are usually quite acceptable to those who work in admissions. There are some obvious problems with them. One cannot be certain that the applicant wrote the statement if it was obtained through the mail. Second, while written statements are readily obtained, reading them carefully is "sheer drudgery," in the words of one reviewer (Davis, 1973). Consequently, they are often not read carefully, and rarely are they scored or coded in any systematic way. As psychological instruments, written statements tend to be viewed with disdain by behaviorists, and one reviewer (Annis, 1967) cites the rise of behaviorism as a reason for the decline in use of autobiographies in psychology. As currently used in admissions, personal statements are viewed as useful and acceptable primarily as an indication of writing proficiency. However, as measures of writing proficiency, written statements are much less reliable than objective tests of writing skill.

Assessment centers

Assessment center techniques require that examinees come to a central location, where they can be assessed in a face-to-face situation. They can also engage in the group activities required for some procedures. When assessment is conducted in this way, the techniques used are essentially those used in settings normally referred to as assessment centers. Assessment centers have been most often used in government, military, and business organizations for the selection of high-level employees. An early example of the assessment center was featured in the activities of the Office of Strategic Services Assessment Staff (1948) during World War II. In more recent years assessment center methods have been used to simulate problems of individuals in organizational life in business and government. Several techniques are commonly used, including the Leaderless Group Discussion (LGD, discussed previously), in-basket exercises which simulate a common aspect of organizational life (Frederiksen, 1957, 1962; Crooks, 1971), peer ratings, paper-and-pencil tests, projective tests, and business games.

Several studies have demonstrated the utility of assessment centers in business contexts. Wollowick and McNamara (1969) studied 94 lower-level and middle managers who participated in an assessment program and later were evaluated with respect to managerial responsibilities attained. The overall rating of four observers in the assessment center were found to correlate significantly (.37) with the position-level criterion. However, a much higher correlation (.62) was obtained by an empirical combination of several scores and ratings. Byham (1970) summarized the results of several studies correlating assessment center evaluation with subsequent

supervisor ratings and reported validities from .27 to .64. Finkle (1976) described several studies in which correlations in the range .10–.74 were obtained. See Byham and Bobin (1972), Cohen, Moses, and Byham (1974), and Moses and Byham (1977) for further validity information. It is clear from these studies that the usefulness of assessment centers varies considerably.

Assessment centers are scheduled for periods of one day to a week and typically involve 12 assessees (divided into two groups of six for certain exercises) and several assessors. Obviously, the cost of such centers is high, and large-scale implementation is usually precluded. For these reasons, any use of assessment centers in admissions to higher education would of necessity be restricted to special cases of applicants.

Minicourses

The minicourse is in some ways an extension of the assessment center concept and in some ways an extension of the work sample concept. Applicants may be engaged for relatively long periods of time in an activity very closely resembling the actual work for which they are being assessed. One example is the summer programs of the Council for Legal Education Opportunity (CLEO), in which disadvantaged law school applicants study law for a summer and are then recommended or not recommended to law schools for admission. Other minicourses have been studied in industry (for example, a study reported by Reilly and Manese (1979) of telephone company switching technicians).

Combined procedures

Up to this point the discussion of procedures has focused on single procedures one at a time. Usually, however, more than one admissions procedure is used and combined either in a holistic judgmental sense or combined statistically. When an admissions officer reads through the folder of an applicant and then makes some kind of rating or several ratings, all the information available is being combined into a single judgment. The same is true of judgments made by an admissions panel. Much of the same information can also be combined in a statistical prediction of some outcome of importance such as academic performance in college.

Lavin (1965) noted the difficulty in summarizing studies of multiple prediction. Often several different procedures are used to predict different outcomes, and the methods of describing how the predictors contribute to prediction vary somewhat. It is thus difficult to identify a simple means of summarizing what the studies have found. As in previous sections of this review, however, it seems appropriate to separate academic and nonacademic outcomes as much as possible. As explained before, this has not been a simple separation because more often than not the nonacademic outcomes contain academic outcomes to some degree. Nevertheless, a judgment has been made where possible and outcomes classified as either academic or nonacademic. Examples of nonacademic outcomes include faculty ratings of medical students and resident interns on clinical performance, scientists' income, creative achievement, peer ratings, and extracurricular activities in college. Also included in nonacademic outcomes were third- and fourth-year medical school grades, since these are most often based on clinical rather than academic performance.

As a second means of summarization two basic questions were asked: (1) Of all the procedures used, which was the best predictor for a given outcome in a given sample? (2) When all predictors were combined statistically, what magnitude of relationship was attained for the prediction of the outcome of interest?

Multiple prediction of academic outcomes

Table 17 summarizes a number of studies in which several procedures were used to predict an academic outcome. In some cases the procedures have been combined statistically to produce a multiple correlation, but in other studies no multiple correlation was reported. Usually, however, it was possible to determine which of the predictors correlated best with the outcome studied. Table 17 indicates the outcome predicted, the predictors used, the best predictor, the correlation between the best predictor and the outcome, and the multiple correlation between all predictors combined and the outcome.

Table 17. Multiple prediction of academic outcomes

Study	Outcome predicted	Predictors used	Sample	Best predictor	Correlations: Best r	Correlations: R
Aiken (1964)	Freshman GPA	Grades, tests, biodata	1	Biodata	.57	–
Ayers & Rohr (1974)	Teacher program GPA	Grades, tests, personality, interests	1	Grades	.81	–
			2	Grades	.78	–
Brown & Dubois (1964)	Freshman GPA	Grades, tests, biodata, personality	1	Personality	.56	.67
			2	Personality	.49	.62
Calkins et al. (1980)	2-yr. examination	Grades, tests, biodata, interviews, recs.	1	Tests	.28	–
			2	Tests	.54	–
	2-yr. GPA		1	Grades	.32	–
			2	Tests	.51	–
Creager (1965)	Doctorate attain.	Grades, tests, recs.	(9)	Tests	.38	.39
Donnan (1968)	Freshman GPA	Grades, tests, personality	1	Grades	.43	.63
			2	Grades	.48	.59
	Sophomore GPA		1	Grades	.42	.62
Dunham (1965)	Freshman GPA	Grades, tests, personality	1	Grades	.44	.67
Gough & Hall (1975)	Academic factor, medical school	Grades, tests, interviews, personality	1	Grades	.39	.43
			2	Grades	.36	.44
			3	Grades	.36	.47
Gough, Hall, & Harris (1963)	Med. sch. GPA, yr. 1	Grades, tests, interview	(14)	Grades, tests	.28	–
	Med. sch. GPA, yr. 2			Grades	.22	–
Harrington (1969)	Freshman GPA	Grades, tests, interests, biodata	1	Grades	.43	.53
		Grades, tests, biodata	2	Tests	.58	.66
Himmelstein (1965)	Freshman GPA	Tests, personality	1	Tests	.60	–
Hoffman et al. (1963)	Med. sch. GPA, yr. 1	Grades, tests, interviews	1	Grades	–	–
	Med. sch. GPA, yr. 2	Grades, tests, interviews		Grades	–	–
Korman et al. (1968)	Academic ach. factor, medical school	Grades, tests, interviews, personality, biodata	1	Tests	.48	–
Lindsay & Althouse (1969)	Freshman GPA	Grades, tests, interests	1	Grades	.38	.45
			2	Tests	.48	.55
Malloy & Ivanoff (1964)	Sophomore GPA	Grades, tests, biodata, personality	1	Biodata	.56	.69
			2	Grades	.56	.69
Nichols (1966)	Freshman GPA	Grades, tests, interests, personality	1	Personality	.36	–
			2	Grades	.39	–

Table 17 continued

Study	Outcome predicted	Predictors used	Sample	Best predictor	Correlations Best r	R
Payne, Rapley, & Wells (1973)	Freshman GPA	Grades, tests, biodata	1	Biodata	.45	.56
			2	Biodata	.41	.53
Reilly (1971)	1st-yr. law GPA	Grades, tests, recs.	1	Grades	.48	.72
Richards & Casey (1975)	Freshman GPA	Grades, tests, personality, creativity	1	Tests	.45	.62
Richards et al. (1974)	Academic status, medical school	Grades, tests, interviews, personality, biodata, interests	1	Grades	.27	.43
Richards & Taylor (1961)	Med. sch. GPA, yr. 1	Grades, tests, interviews	1	Grades, tests	.24	.40
	Med. sch. GPA, yr. 2			Grades	.39	—
	Med. sch. GPA, yr. 1	Grades, tests, recs.	2	Grades	.36	—
	Med. sch. GPA, yr. 2			Grades	.42	—
Stroup (1970)	Freshman GPA	Tests, personality	1	Tests	.47	.60
			2	Grades	.43	.54
Webb (1960)	Freshman GPA	Grades, tests, biodata	1	Grades	.48	.66
			2	Grades	.45	.60
Willingham (1965)	Freshman GPA	Grades, tests, biodata	1	Grades	.49	.67
Median correlations					*.44*	*.60*

Note: When the sample is given in parentheses the figure given is the number of samples studied, and the correlations reported are medians. The term "tests" is used to represent traditional admissions tests.

A number of observations can be made from Table 17. As one would expect, the best predictor of academic outcomes is previous grades. Of the 45 best predictors indicated in Table 17, 26 are previous grades and 10 are tests of some type. Rarely do other kinds of measures do as well in the prediction of academic outcomes. In the study of Brown and Dubois (1964) the best predictor shown is personality measures. Actually, the "personality" measure that predicted so well in that particular study was part of a "biographical inventory" and, specifically, was an item that asked students to provide the lowest GPA that would be acceptable to them in college. While not a formal measure of personality, the response to this kind of question is suggestive of motivation, drive, and attitude normally classified as personality. But, in a sense, such a question is also a proxy for grades, since students with good previous grades will expect to receive good grades in college. Personality was also the best predictor in Nichols's (1966) study, in which an objective behavior inventory was used to develop a scale for the prediction of freshman GPA. In none of these studies or samples were interviews, recommendations, or interest measures the best predictors.

The median correlations shown at the bottom

Table 18. Multiple prediction of nonacademic outcomes

Study	Outcome predicted	Predictors used	Sample	Best predictor	Correlations: Best r	Correlations: R
Anastasi et al. (1960)	College achievement	Grades, tests, biodata	1	Biodata	.55 (phi)	–
Antastasi &Schaefer (1969)	HS art, achievement		1	Biodata	.34 (phi)	–
Astin (1964)	Persistence	Biodata, personality	1	Personality	.18	.29
			2	Personality	.28	.44
Calkins et al. (1980)	2-yr. clinical perf.	Grades, tests, biodata, interviews, recs.	1	Grades, tests	.17	–
			2	Tests	.45	–
	2-yr. prog. status		1	Tests	.18	–
			2	Tests	.42	–
Creager & Harmon (1966)	Scientist income	Grades, tests, recs.	(10)	Tests	.20	
	Scientist product.			Tests	.27	
	Scientist citations			Tests	.27	
	On-the-job ratings			Recs.	.35	
Donnan (1968)	Persistence, 1 yr.	Grades, tests, personality	1	Personality	.23	.31
	Persistence, 1 yr.		2	Grades	.18	.22
	Persistence, 2 yr.		1	Tests	.28	.38
Gough & Hall (1975)	Clinical perf., medical school	Grades, tests, interviews, personality	1	Personality	.13	.17
			2	Personality	.11	.11
			3	Personality	.25	–
Gough, Hall, & Harris (1963)	Med. sch. GPA, yr. 3	Grades, tests, interviews	(14)	Grades	.17	–
	Med. sch. GPA, yr. 4			Tests	.10	–
Gough, Hall, & Harris (1964)	Faculty Factor I	Grades, tests, interviews	1	Tests	.09	–
	Faculty Factor II			Interviews	.09	–
	Faculty Factor III			Interviews	.24	–
	Faculty Factor I′			Interviews	.13	–
	Faculty Factor II′			Interviews	.21	–
	Peer Factor I			Grades	.24	–
	Peer Factor II			Grades	.08	–
	Peer Factor III			Tests	.10	–
Heath (1977)	Adult maturity	Grades, tests, personality	1	Personality	–	–
Hoffman et al. (1963)	Med. sch. GPA, yr. 3	Grades, tests, interviews	1	Personality		
Keck et al. (1979)	Resident evaluations	Grades, tests, interviews, interests, personality	1	Tests	.39	–

Table 18 continued

Study	Outcome predicted	Predictors used	Sample	Best predictor	Correlations: Best r	Correlations: R
Kegel-Flom (1975)	Intern perf. (supv.) Intern perf. (peer)	Grades, tests, interviews, interests, personality	1	Interests Personality	.25 .39	– –
Korman et al. (1968)	Internship success Peer esteem Humanism	Grades, tests, biodata, personality, interviews	1	Biodata Biodata Biodata	.31 .47 .37	– – –
Nichols (1966)	Extracurricular activities	Grades, tests, interests, personality	1 2 3 4	Personality Tests Personality Personality	.30 .16 .23 .17	– – – –
Nichols & Holland (1963)	Nonacademic composite of college achievement	Grades, tests, biodata, interests, personality, creativity, parent ratings, self-ratings	1 2	Biodata Biodata	.32 .26	– –
Richards & Lutz (1968)	College accomplishment in specific areas	Grades, tests, biodata	1	Biodata	.25	–
Schaefer (1969)	HS writing ach.	Biodata, personality	1	Biodata	.64	–
Scott & Castles (1975)	HS talented accomplishments	Biodata, personality, tests	1 2	Biodata Biodata	.33 .67	– –
Median correlations					.25	.29

of Table 17 are indicative of the magnitude of relationship usually obtained in the prediction of academic outcomes, and these medians are comparable to those reported elsewhere for this kind of prediction.

Multiple prediction of nonacademic outcomes

Table 18 summarizes those studies reporting predictions of nonacademic outcomes using a number of different procedures. Interestingly, traditional admissions tests were most often (in 12 cases) the best predictor of these nonacademic outcomes; personality and biodata ran a close second (11 cases each). Grades were the best predictor in 9 cases, interviews best in 5 cases, and recommendations in none. Most interesting in Table 18, however, is the general magnitude of prediction of nonacademic outcomes. The median correlation for the best predictors is only .25 and that for multiple prediction only .29 (even though this last figure is based on only seven correlations). While the median correlations are low and thus suggest that nonacademic outcomes are not very predictable, the range of the correlations is quite high. The studies of Anastasi et al. (1960), Schaefer (1969), and Scott and Castles (1975) achieved predictive correlations of .55 or better.

Since all these studies used carefully developed criterion measures, they suggest that good predictions of nonacademic outcomes can be made if the criteria are valid reliable measures.

There have been two studies using multiple procedures that do not fit well into comparisons of correlations but that are still of special interest. Nicholson (1970) made a long-range study of Brown University students from the time they were freshmen until 15 years later. Biodata, recommendations, interests, and traditional aptitude tests were all related to occupational level attained and to alumni judgments of peers over this period. With the exception of the interest measure (SVIB), a mathematical aptitude test score (SAT-M), and one of two biodata measures obtained at the time of application for admission to Brown, all admissions variables were significantly related to occupational level attained 15 years after graduation—or almost 20 years after the measures were obtained. The biodata measure that did relate to long-term occupational achievement was derived from responses to a survey of high school leadership and athletic and social participation. The biodata measure that did not relate to long-term occupational achievement was a measure of high school literary, oratorical, musical, artistic, and scientific participation. Less consistent results were obtained where the peer judgments were used as the criterion of long-term success.

A study by Heath (1977) also used multiple measures to predict long-term outcomes. Traditional aptitude test scores, grades, and personality measures were used to predict emotional maturity (as judged by friends, peers, and others and as measured by personality tests such as the MMPI) of Haverford College graduates who were in their thirties and employed. Although very little evidence was presented, it was argued that only personality measures and judgments of maturity predicted the long-term outcomes studied.

Summary and conclusions

In this review several different procedures used for the assessment of student characteristics, as well as other procedures with potential for use, have been considered. Evidence on the utility of these procedures has been assembled from a wide range of settings because the available evidence from any one setting was at times limited. Studies of admissions procedures for colleges, graduate schools, law schools, business schools, and medical schools were included, as were studies in industrial, business, military, and governmental settings. Medical school admissions research was found to be a particularly rich source of evidence. Much less use was made of research from employment settings, but in some cases this was the only source of relevant information.

The review has generally moved from the most-common to the least-common procedures, with emphasis on the former. Thus, the review began with the application blank and its expanded version, the biographical inventory. Recommendations, interviews, interest measures, and personality measures were then considered in that order. Finally, a number of other procedures, often discussed but less-often used or researched in admissions per se, were described briefly.

The focus of the review has been on the reliability, validity, and acceptability of these procedures. Acceptability includes a number of considerations including economic feasibility, practicality, workability, fakability, equity and privacy. In analyses of predictive validity special attention was given to the problem of criteria, and, where possible, academic and nonacademic outcomes were treated separately.

Reliability

Except for the more formal measures reviewed—interest measures and personality measures—the reliability of procedures used at present in admissions work has not been well researched. The few available studies of the reliability of interviews and recommendations present a rather bleak picture. Biographical inventories are an enigma of a sort because test-retest evidence is the only appropriate indication of their reliability. Internal consistency estimates are confounded by the heterogeneity of such inventories. More attention has been given to the examination of reliability in the newer procedures, perhaps because they are being more carefully developed prior to widespread use.

Validity

Most of the available evidence on the validity of admissions procedures is of the predictive type; this emphasis is indicative of a traditionally narrow view of the admissions process and of the goals of higher education. For those procedures studied often enough to allow for computations of median predictive validities, Table 19 summarizes

Table 19. Predictive validity comparisons of selected procedures

Procedures	Median predictive correlations	
	Academic outcomes	Nonacademic outcomes
Biodata	.43	.35
Recommendations	.20	.18
Interviews	.15	.10
Interest measures	.12	.26
Personality measures	.28	.13
Combined procedures[a]	.60	.29

a. Includes grades and test scores.

the results for the prediction of academic and nonacademic outcomes. This summary would indicate that biodata offer some promise for use in admissions. With the exception of interest measures, the other procedures seem less promising—especially where academic outcomes are better predicted than nonacademic outcomes. Traditional procedures usually exceed these levels of predictive accuracy. Combined procedures are of course usually better predictors than any single procedure, but economic and other considerations limit the number of measures that can be used in practice.

An especially interesting phenomenon observable from Table 19 is the contrast of combined procedures as presented here in the sense of statistical or actuarial combinations and combinations of information made in a clinical sense (as in interviews and recommendations). The higher correlations for the statistical combinations, whether in the prediction of academic or nonacademic outcomes, buttress the arguments that have been made so eloquently by Hoffman (1960), Dawes (1971), and Dawes and Corrigan (1974). This argument is essentially that a linear combination of data in a regression equation consistently results in more reliable assessments and more valid predictions than do combinations of the same data when made in the clinical sense of holistic judgments either by individuals or by committees. Most of the interviewer and referee judgments included in this review and summarized in Table 19 were made with access to ancillary data on the individuals being judged.

Examination of the issue of incremental validity showed that studies have demonstrated statistically significant increases in prediction beyond grades and test scores for biodata, interest measures, and personality measures. Often, however, adequate cross-validation was not conducted, and consequently the demonstrations are not altogether convincing. One doubts, moreover, whether the small increments shown offer sufficient justification in and of themselves for the use of these procedures. Better arguments can be made, it seems, for the use of biodata and interests as sources of diversity, as means for promoting exchanges of information useful to both students and institutions, and as ways in which broader objectives of students and institutions can be realized.

Acceptability

Perhaps as important as reliability or validity is whether the procedures reviewed are acceptable to the institutions and to the students who use them. As explained in the introduction to this report, acceptability includes a wide range of considerations important in admissions. Not all the procedures reviewed would satisfy these acceptability considerations equally well, and some procedures would in fact tend to work against these kinds of interests.

As for biographical data, it is clear that one cannot summarize their acceptability as a part of admissions procedures unless the type of biographical data is specified. Moreover, whether a particular type of biographical data is acceptable depends on the outcome being considered. An applicant's accomplishments and activities in the past certainly would appear to be acceptable to

most institutions and to students as information relevant to decisions in admissions. However, they may not predict academic performance. When these accomplishments and activities are supplied by the applicant without any verification of their accuracy from other persons, acceptability becomes more of a question. The usual demographic data—sex, ethnicity, state of residence, and the like—seem quite acceptable if solicited in the proper manner. An institution may wish to use such information to provide balance and diversity and to plan for needed facilities. Some information may be acceptable for collection during admissions but not acceptable for making admissions decisions. Marital status and parents' income, for example, would not be acceptable for use in making decisions on acceptance, even though they would be useful for other purposes. Whether it is acceptable to use empirical keying techniques with biographical inventories depends on what the content of the inventory is. The content must be relevant to what is expected of a student in a given academic program if it is to receive some kind of weighting in an admissions index based on a biographical inventory. However, the most acceptable kind of biographical information may be that which is simply used as is without any kind of empirical weighting scheme. Biographical data on past activities and accomplishments may be especially acceptable because of evidence indicating that group differences may be minimal.

Interviews were seen in this review as practical, workable, and acceptable to a broad public. As an aid in public relations, they also seem to be important with respect to enrollment concerns. And as a source of information for students about the program or institution they are applying to, the interview is of great importance. There are two principal problems of acceptability with the interview, both primarily economic in nature but both partially problems of equity. All applicants do not live in close proximity to the institution applied to; as a result, the student who lives at a distance is at an economic disadvantage if appearing for an interview enhances chances of admission. Even if the interview is not used in admissions decisions, the distant applicant is at an informational disadvantage. The second problem is that it is costly to the institution when large numbers of applicants come for interviews.

Recommendations have a number of acceptability problems. First, they require considerable time and commitment from those who prepare them. Second, they are a possible source of offense for certain groups, particularly women, who may at times be inappropriately represented by the referee. Third, the recommendation requirement is an additional hurdle for the applicant in the sense that it is an additional task that must be completed. In some application situations, the applicant may not wish to reveal his or her plans; a requirement to do so thus constitutes a kind of invasion of privacy. Whether recommendations have adverse impact on protected groups is unknown, but it is a possibility.

Interest measures can, of course, be faked. The more formal interest inventories could be expensive to maintain if implemented in an admissions program. On the other hand, formal measures of interests might be perceived as a positive experience by many applicants. The possibility of faked responses could be less important however, if interest information were used primarily for guidance and for creating institutional diversity. If the time required to complete the more formal interest inventories or the expense of maintaining them are seen as problems by an institution, research evidence would suggest that brief ratings by applicants of their interests can serve as a useful substitute for the lengthier inventories. Such brief ratings of interests could easily be included as part of the application form.

Personality measures are of doubtful acceptability for regular use in admissions. Many consider them to be an invasion of privacy, and, ac-

cordingly, widespread public acceptance cannot be anticipated. Moreover, they can be faked and may not even be taken seriously by applicants. Applicants may see them as an obstacle they would rather not experience and thus prefer to apply to other institutions where they are not required. Finally, the ethical issue seems almost insurmountable. If a given personality type is known to perform best in a particular academic program, is it ethical to prefer that personality type over another?

Summarization of considerations

As a heuristic device it is useful to attempt a summarization of the reliability, validity, and acceptability considerations used in reviewing the various procedures covered. There is a danger in viewing such a summarization as evaluative, however, because evaluative concerns would necessarily deal with very specific instruments—not procedures grouped into broad categories, as has been done in this book. And, additionally, evaluative considerations may be contradictory within the same instrument because different types of content may be involved. This last complication is certainly the case for biographical data, since some items may be content-relevant for the purpose intended, while other items may not be. Nevertheless, Table 20 attempts a heuristic summary of the review considerations for the five procedures that have been the focus of this review.

Table 20 is, of course, a gross oversimplification of a great many complexities. Consequently, it must be viewed with some caution. On the other hand, it presents a capsule description of the five procedures in terms of the three major considerations of the review. Note that two of the procedures—biographical data and interest measures—stand out because they have +'s in all three columns. As indicated in Table 20, it does not follow that all types of biographical information or all interest measurements are equally promising or that these two procedures have no problems. Nor does it follow that all recommendations, interviews, and personality measures are of questionable value. But as a means of directing research and policy on admissions, Table 20 is useful in suggesting where the most fruitful effort may be expended. As a result of the information collected during this review, it does not seem likely that the problems of recommendations, interviews, and personality measures are likely to be easily resolved. On the other hand, biographical data and interest measures of certain types appear to offer considerable promise for use in admissions work.

Table 20. A heuristic summary

Procedure	Consideration		
	Reliability	Validity	Acceptability
Biographical data	+	+	+
Recommendations	?	?	?
Interviews	−	−	+
Interest measures	+	+	+
Personality measures	+	−	−

Note: + indicates that the evidence is generally positive but not always.

− indicates that the evidence is generally negative but not always.

? indicates that the evidence is either insufficient or mixed.

A similar type of heuristic summary might be used for the other procedures described briefly in this review. However, the lack of experience with these procedures in admissions settings argues against such a summarization. But it could be that some of the newer procedures offer more promise than the five summarized here and will eventually surface as solutions to important admissions problems.

Research recommendations

Having completed a review as extensive as this one, one would expect that some recommendations might be forthcoming with respect to re-

search that would further the objectives of interest. With so many imperfect procedures, there is no lack of places where efforts could be made. And the many new thoughts on assessment and new ways of thinking about admissions offer other avenues of research that seem quite interesting. Nonetheless, there exist exemplary precedents for continuing with the kinds of research where the most experience has been obtained. Of the five types of procedures emphasized, two stand out as offering promise in terms of reliability, validity, and acceptability. The following specific research activities seem worthy of consideration.

1. *A comprehensive questionnaire.* While biographical data and interests seem to offer solutions to some contemporary admissions problems, they both confront practical difficulties that may be helped through research. There is a need to find acceptable means through which self-reported biographical information can be verified and thus made more credible. And careful review of the content of biographical inventories as well as application blanks needs to be conducted to assure that the questions asked are the most appropriate for the purposes intended. Formal interest assessment may be too complex and time consuming for use in admissions. If that is so, then some evidence suggests that very brief indications of interests are almost as useful. A research effort could determine if it is feasible to integrate both of these into a single comprehensive questionnaire.

2. *Improved reference recommendation procedures.* With the low reliability and validity of the recommendation procedures used at present and with increasing concern about their acceptability, there is a need to determine ways in which these procedures could be redirected. One possible approach is to integrate an inventory of accomplishments and interests with the recommendation. The referee would then verify or augment the inventory.

3. *Improved interview procedures.* Although the interview as used at present is generally acceptable to both institutions and students, perhaps it should not be. Research directed toward improved efficiency in the interview as well as toward increased reliability and validity could be of considerable value to those who work in admissions. Since structured interviews are the usual outcome of efforts to improve their reliability and validity, and since a structured interview is tantamount to a biographical inventory, the integration of the inventory of accomplishments and interests with the interview (as well as with the recommendation) is one suggestion for improvement. Research would be necessary, however, to determine how such an integration could be achieved.

4. *Formal interest assessment.* It is not entirely clear that the more formal measures of interest need to be replaced with a brief assessment of interests for the sake of economy, acceptability, etc. It may be that applicants to educational programs would very much enjoy an extensive exploration of their interests and would find that such an exploration yielded great benefits to them in their educational and career planning.

5. *Paramorphic representations of judgments.* The evidence of this review strongly suggests that neither interviewers nor referees make consistent judgments about applications to educational programs in higher education. There is one line of research that suggests how improvements in interviewer and referee judgments may be made more consistent through what has been termed a "paramorphic" representation of more specific kinds of judgments—i.e., a statistical combination of judgments and other ancillary information available. A review of this area of research could be useful, as could an attempt to implement some of the ideas from it into admissions work.

6. *Reliability of admissions procedures.* Commonly used admissions procedures such as interviews and recommendations have low reliabilities. But very little research has been conducted either

to examine reliability systematically or to determine ways in which it might be improved.

7. *Criterion development.* One of the major reasons that nonacademic outcomes are so poorly predicted is that the means available for assessing them tend to be unreliable. There is a need to improve criterion assessment so that nonacademic admissions variables can be better validated.

8. *New developments.* The many new procedures now being researched and developed may offer solutions to some admissions problems. These need to be closely watched as research continues.

References

Aiken, L. R., Jr. The prediction of academic success and early attrition by means of a multiple-choice biographical inventory. *American Educational Research Journal*, 1964, *1*(2), 127–135.

Amir, Y.; Kovarsky, Y.; and Sharan, S. Peer nominations as a predictor of multistage promotions in a ramified organization. *Journal of Applied Psychology*, 1970, *54*, 462–469.

Anastasi, A. *Psychological testing*, 4th ed. New York: Macmillan Publishing Co., Inc., 1976.

Anastasi, A.; Meade, M. J.; and Schneiders, A. A. *The validation of a biographical inventory as a predictor of college success: Development and validation of the scoring key*. New York: College Entrance Examination Board, 1960.

Anastasi, A., and Schaefer, C. E. Biographical correlates of artistic and literary creativity in adolescent girls. *Journal of Applied Psychology*, 1969, *53*, 267–273.

Annis, A. P. The autobiography: Its uses and value in professional psychology. *Journal of Counseling Psychology*, 1967, *14*(1), 9–17.

Argyris, C. *Interpersonal competence and organizational effectiveness*. Homewood, Illinois: Irwin, 1962.

Asher, J. J. Reliability of a novel format for the selection interview. *Psychological Reports*, 1970a, *26*, 451–456.

Asher, J. J. How the applicant's appearance affects the reliability and validity of the interview. *Educational and Psychological Measurement*, 1970b, *30*(3), 687–695.

Asher, J. J. Q by Q interview as a predictor of success in second language learning. *Psychological Reports*, 1971, *29*, 331–337.

Asher, J. J., and Sciarrino, J. A. Realistic work sample tests: A review. *Personnel Psychology*, 1974, *17*(4), 519–523.

Atkinson, J. W. *Motives in fantasy, action and society: A method of assessment and study*. Princeton: D. Van Nostrand Company, 1958.

Ayers, J. B., and Rohr, M. E. Relationship of selected variables to success in a teacher preparation program. *Educational and Psychological Measurement*, 1974, *34*, 933–937.

Baehr, M., and Williams, G. B. Underlying dimensions of personal background data and their relationship to occupational classification. *Journal of Applied Psychology*, 1967, *51*, 481–490.

Baird, L. L. *Family income and the characteristics of college-bound students*. ACT Research Report No. 17. Iowa City: American College Testing Program, 1967.

Baird, L. L. Prediction of accomplishment in college: A study of achievement. *Journal of Counseling Psychology*, 1969a, *16*(3), 246–253.

Baird, L. L. Factors in the continuance of accomplishment from high school to college. *Measurement and Evaluation in Guidance*, 1969b, *2*(1), 5–18.

Baird, L. L. *Patterns of educational aspiration*. ACT Research Report No. 32. Iowa City: American College Testing Program, 1969c.

Baird, L. L. Prediction of academic and non-academic achievement in two-year colleges from the ACT assessment. *Educational and Psychological Measurement*, 1969d, *29*, 421–430.

Baird, L. L. *Using self-reports to predict student performance*. Research Monograph No. 7. New York: College Entrance Examination Board, 1976.

Baird, L. L. *Development of an inventory of documented accomplishments for graduate admissions*. GRE Board Research Report GREB No. 77-3R. Princeton: Educational Testing Service, 1979.

Baird, L. L., and Richards, J. M., Jr. *The effects of selecting college students by various kinds of high school achievement*. ACT Research Report No. 23. Iowa City: American College Testing Program, 1968.

Banreti-Fuchs, K. M. Attitudinal, situational, and mental health correlates of academic achievement at the undergraduate university level. *British Journal of Educational Psychology*, 1975, *45*, 227–231.

Banreti-Fuchs, K. M., and Meadows, W. M. Interest, mental health, and attitudinal correlates of academic achievement

among university students. *British Journal of Educational Psychology*, 1976, *46*, 212–219.

Banta, G. W. *A comparison of the leaderless group discussion and the individual interview techniques for the selection of student-orientation assistants.* Unpublished doctoral thesis, University of Tennessee, 1967.

Barrows, H. S., and Tamblyn, R. M. An evaluation of problem-based learning in small groups utilizing a simulated patient. *Journal of Medical Education*, 1976, *51*, 52–54.

Bayer, A. E., and Boruch, R. F. The black student in American colleges. ACE *Research Reports*, *4*(2). Washington, D.C.: American Council on Education, 1969.

Belcastro, F. P. Use of selected factors as predictors of success in completing a secondary teacher preparation program. *Educational and Psychological Measurement*, 1975, *35*, 957–962.

Best, A. M., and Samph, T. An experimental paper-and-pencil examination of medical student interpersonal skills. Paper presented at the annual meeting of the American Educational Research Association, San Francisco, April 1979.

Brown, F. G., and Dubois, T. E. Correlates of academic success for high-quality freshman men. *Personnel and Guidance Journal*, 1964, *42*, 603–607.

Browning, R. C. Validity of reference ratings from previous employers. *Personnel Psychology*, 1968, *21*(3), 389–393.

Plumback, G. B. A note on criterion contamination in the validation of biographical data. *Educational and Psychological Measurement*, 1969, *29*, 439–443.

Bryson, J. B. The dimensions of early human experience. In W. A. Owens (chair.), *Developmental implications of biographical data.* Symposium presented at the meeting of the Southern Society for Philosophy and Psychology, Miami, April 1969.

Burgess, M. M.; Calkins, V.; and Richards, J. M. The structured interview: A selection device. *Psychological Reports*, 1972, *31*, 867–877.

Buros, O. K. *The sixth mental measurements yearbook.* Highland Park, New Jersey: The Gryphon Press, 1965.

Buros, O. K. *The seventh mental measurements yearbook.* Highland Park, New Jersey: The Gryphon Press, 1972.

Byham, W. C. Assessment centers for spotting future managers. *Harvard Business Review*, 1970, *48*(4), 150–160.

Byham, W. C., and Bobin, D. (eds.). *Alternatives to paper and pencil testing.* Proceedings of a conference sponsored by the Graduate School of Business, University of Pittsburgh. Pittsburgh: University of Pittsburgh, 1972.

Calkins, E. V.; Richards, J. M., Jr.; McCanse, A.; Burgess, M. M.; and Willoughby, T. L. Impact on admission to a school of medicine of an innovation in selection procedures. *Psychological Reports*, 1974, *35*, 1135–1142.

Calkins, E. V.; Willoughby, T. L.; and Arnold, L. M. *Predictors of performance of minority students in the first two years of a BA/MD program.* Unpublished report, University of Missouri (Kansas City) School of Medicine, 1980.

Campbell, D. P. *Manual for the Strong Vocational Interest Blanks for men and women.* Stanford, California: Stanford University Press, 1966.

Carlson, R. E. Selection interview decisions: The relative influence of appearance and factual written information on an interviewer's final rating. *Journal of Applied Psychology*, 1967, *51*(6), 461–468.

Carlson, R. E. The current status of judgmental techniques in industry. In W. C. Byham and D. Bobin (eds.), *Alternatives to paper and pencil testing: Proceedings of a conference.* Pittsburgh: Graduate School of Business, University of Pittsburgh, 1972.

Carlson, R. E., and Mayfield, E. C. Evaluating interview and employment application data. *Personnel Psychology*, 1967, *20*, 441–460.

Cascio, W. F. Accuracy of verifiable biographical information blank responses. *Journal of Applied Psychology*, 1975, *60*(6), 767–769.

Cascio, W. F., and Phillips, N. F. Performance testing: A rose among thorns? *Personnel Psychology*, 1979, *32*, 751–766.

Cassens, F. P. *Cross cultural dimensions of executive life history antecedents.* Greensboro, North Carolina: Creativity Research Institute, Richardson Foundation, Inc., February 1966.

Chaney, F. B., and Owens, W. A. Life history antecedents of sales, research, and general engineering interests. *Journal of Applied Psychology*, 1964, *48*, 101–105.

Char, W. F.; McDermott, J. F., Jr.; Haning, W. F., III; and Hansen, M. J. Interviewing, motivation, and clinical judgment. *Journal of Medical Education*, 1975, *50*, 192–194.

Chen, M. K.; Podshadley, D. W.; and Shrock, J. G. A factorial study of some psychological, vocational interest, and mental ability variables as predictors of success in dental school. *Journal of Applied Psychology*, 1967, *51*(3), 236–241.

Clendenen, D. M. (review). In O. K. Buros, *The seventh mental measurements yearbook*, Volume II, pp. 1471–1473. Highland Park, New Jersey: The Gryphon Press, 1972.

Coffman, W. E. On the validity of essay tests of achievement. *Journal of Educational Measurement*, 1966, *3*(2), 151–156.

Cohen, B.; Moses, J. L.; and Byham, W. C. *The validity of assessment centers.* Pittsburgh: Development Dimensions Press, 1974.

Creager, J. A. *Predicting doctorate attainment with* GRE *and other variables.* Technical Report No. 25. Office of Scientific Personnel, National Academy of Sciences – National Research Council, Washington, D.C. November 16, 1965.

Creager, J. A., and Harmon, L. R. *On-the-job validation of*

selection variables. Technical Report No. 26 (abstract). Office of Scientific Personnel, National Academy of Sciences – National Research Council, Washington, D.C. April 15, 1966.

Crooks, L. A. *The in-basket study: A pilot study of* MBA *candidate performance on a test of administrative skills as related to selection and achievement in graduate business school.* ATGSB Brief No. 4. Princeton: Educational Testing Service, 1971.

Crooks, L. A. *The selection and development of performance measures for assessment center programs* (RM 74-6). Princeton: Educational Testing Service, 1974.

Cuca, J. M.; Sakakeeny, L. A.; and Johnson, D. G. *The medical school admissions process: A review of the literature, 1955–76*. Washington, D.C.: Association of American Medical Colleges, 1976.

Davis, E. E. An analysis of the autobiographical statements of adolescents. *Measurement and Evaluation in Guidance*, 1973, *6*(2), 117–121.

Dawes, R. M. A case study of graduate admissions: Application of three principles of human decision making. *American Psychologist*, 1971, *26*, 180–188.

Dawes, R. M., and Corrigan, B. Linear models in decision making. *Psychological Bulletin*, 1974, *81*(2), 95–106.

Doering, J. V.; Killup, D. E.; and Fuller, J. L. Can an admissions interview change the character of a dental class? *Journal of Medical Education*, 1979, *54*(9), 735–736.

Dolliver, R. H. Strong Vocational Interest Blank versus expressed vocational interests: A review. *Psychological Bulletin*, 1969, *72*(2), 95–107.

Donnan, H. Personality factors related to college achievement and attrition. *Journal of College Personnel*, 1968, *9*(2), 116–119.

Dunham, R. B. Achievement motivation as predictive of academic performance: A multivariate analysis. *The Journal of Educational Research*, 1973, *67*(2), 70–72.

Dunnette, M. D. Personnel management. *Annual Review of Psychology*, 1962, *13*, 285–314.

Dunnette, M. D. *Handbook of industrial and organizational psychology*. Chicago: Rand McNally, 1976.

Dyer, H. S. *The interview as a measuring device in education*. ERIC TM Report 56. Princeton: Educational Testing Service, 1976.

Edwards, R. C. Personal traits and "success" in schooling and work. *Educational and Psychological Measurement*, 1977, *37*, 125–138.

England, G. W. *Development and use of weighted application blanks*. Dubuque, Iowa: William C. Brown, 1961.

England, G. W., and Paterson, D. G. Selection and placement – the past ten years. In H. G. Heneman, Jr., L. C. Brown, M. K. Chandler, R. K. Parnes, and H. S. Schulz (eds.), *Employment relations research*. New York: Harper, 1960, 43–72.

Entwistle, N. J. Personality and academic attainment. *British Journal of Educational Psychology*, 1972, *42*, 137–51.

Farley, J. Academic recommendations: Males and females as judges and judged. AAUP *Bulletin*, 1978, *64*, 82–85.

Farley, J. A., and Mayfield, E. C. Peer nominations without peers? *Journal of Applied Psychology*, 1976, *61*(1), 109–111.

Feld, S., and Smith, C. P. An evaluation of the objectivity of the method of content analysis. In J. W. Atkinson (ed.), *Motives in fantasy, action, and society*. Princeton: D. Van Nostrand, 1959.

Finkle, R. B. Managerial assessment centers. In M. D. Dunnette, *Handbook of industrial and organizational psychology*. Chicago: Rand McNally, 1976.

Fishman, J. A., and Pasanella, A. K. College admission-selection studies. *Review of Educational Research*, 1960, *30*(4), 298–310.

Frank, L. L., and Hackman, J. R. Effects of interviewer–interviewee similarity on interview objectivity in college admissions interviews. *Journal of Applied Psychology*, 1975, *60*(3), 356–360.

Frederiksen, N. The evaluation of personal and social qualities. In *College Admissions*. New York: College Entrance Examination Board, 1954, 93–105.

Frederiksen, N. Factors in in-basket performance. *Psychological Monographs*, 1962, *76* (22, Whole No. 541).

Frederiksen, N.; Saunders, O. R.; and Wand, B. The in-basket test. *Psychological Monographs*, 1957, *71* (9, Whole No. 438).

Frederiksen, N., and Ward, W. C. Measures for the study of creativity in scientific problem solving. *Applied Psychological Measurement*, 1978, *2*(1), 1–24.

Freeberg, N. E. The biographical information blank as a predictor of student achievement: A review. *Psychological Reports*, 1967, *20*(3), 911–925.

Freeberg, N. E. Relevance of rater–ratee acquaintance in the validity and reliability of ratings. *Journal of Applied Psychology*, 1969, *53*, 518–524.

Gee, H. H., and Cowles, J. (eds.). Appraisal of applicants to medical school. Report of 1956 Teaching Institute. *Journal of Medical Education*, 1957, *32* (10, Part 2), 1–288.

Godshalk, F. I.; Swineford, F.; and Coffman, W. E. *The measurement of writing ability*. Research Monograph No. 6. New York: College Entrance Examination Board, 1966.

Gough, H. G. Nonintellectual factors in the selection and evaluation of medical students. *Journal of Medical Education*, 1967, *42*, 642–650.

Gough, H. G., and Hall, W. B. The prediction of academic and clinical performance in medical school. *Research in Higher Education*, 1975, *3*, 301–314.

Gough, H. G.; Hall, W. B.; and Harris, R. E. Admissions procedures as forecasters of performance in medical training. *Journal of Medical Education*, 1963, *38*(12), 983–998.

Gough, H. G.; Hall, W. B.; and Harris, R. E. Evaluation of performance in medical training. *Journal of Medical Education*, 1964, *39*(12), 679–692.

Grant, D. L., and Bray, D. W. Contributions of the interview to assessment of management potential. *Journal of Applied Psychology*, 1969, *53*, 24–34.

Guilford, J. P. *The nature of human intelligence*. New York: McGraw-Hill, 1967.

Guillemin, J.; Holstrom, L. L.; and Garvin, M. Judging competence: Letters of recommendation for men and women faculties. *School Review*, 1979, *87*(2), 157–170.

Guion, R. M. *Personnel testing*. New York: McGraw-Hill, 1965.

Hackman, J. D., and Taber, T. D. Patterns of undergraduate performance related to success in college. *American Educational Research Journal*, 1979, *16*(2), 117–138.

Haggerty, H. R. *Personnel research for the U.S. Military Academy, 1942–1953*. U.S. Army Personnel Research Board Technical Research Report, 1077, October 1953.

Harmon, L. R. *Fourteen years of research on fellowship selection*. National Academy of Sciences (Publication 1420). Washington, D.C.: National Research Council, 1966.

Harrington, C. Forecasting college performance from biographical data. *Journal of College Student Personnel*, 1969, *10*, 156–160.

Harris, B. M. Admissions screening with leaderless group discussions. *Journal of College Student Personnel*, 1964, 38–41.

Heath, D. H. Academic predictors of adult maturity and competence. *Journal of Higher Education*, XLVIII(6), 1977, 613–632.

Hendricks, M.; Guilford, J. P.; and Hoepfner, R. *Measuring creative social intelligence*. Psychological Laboratory Report No. 42. Los Angeles: University of Southern California, 1969.

Henry, E. R. Conference on the use of biographical data in psychology. *American Psychologist*, 1966, *21*, 247–249.

Hills, J. R. Use of measurement in selection and placement. In R. L. Thorndike (ed.), *Educational Measurement*, 2nd ed. Washington, D.C.: American Council on Education, 1971, 680–732.

Hilton, T. L., and Myers, A. E. Personal background, experience, and school achievement: Investigation of the contribution of questionnaire data to academic prediction. *Journal of Educational Measurement*, 1967, *4*(2), 69–80.

Hilton, T. L., and Rhett, H. *The base-year survey of the National Longitudinal Study of the High School Class of 1972*. Princeton: Educational Testing Service, 1973.

Himelstein, P. Validities and intercorrelations of MMPI subscales predictive of college achievement. *Educational and Psychological Measurement*, 1965, *25*(4), 1125–1128.

Hoffman, E. L.; Wing, C. W., Jr.; and Lief, H. I. Short- and long-term predictions about medical students. *Journal of Medical Education*, 1963, *38*, 852–857.

Hoffman, P. J. The paramorphic representation of clinical judgment. *Psychological Bulletin*, 1960, *57*(2), 116–131.

Holland, J. L. The prediction of college grades from the California Psychological Inventory. *Journal of Educational Research*, 1959, *50*, 135–142.

Hollander, E. P. The reliability of peer nominations under various conditions of administration. *Journal of Applied Psychology*, 1957, *41*(2), 85–90.

Hollander, E. P. Validity of peer nominations in predicting a distant performance criterion. *Journal of Applied Psychology*, 1965, *49*(6), 434–438.

Holmes, D. H., and Hertel, R. Predictions of GPA from application biographies. *Journal of Medical Education*, 1967, *42*, 269.

Holmes, D. S. Conscious self-appraisal of achievement motivation: The self-peer rank method revisited. *Journal of Consulting and Clinical Psychology*, 1971, *36*(1), 23–26.

Holmes, D. S., and Tyler, J. D. Direct versus projective measurement of achievement motivation. *Journal of Consulting and Clinical Psychology*, 1968, *32*(6), 712–717.

Institute for Behavioral Research in Creativity. *Manual for Alpha biographical inventory*. Greensboro, North Carolina: Predictions Press, 1968.

Jackson, G. G., and Kellow, W. F. An experiment with the group interview in the selection of medical students. *Journal of Medical Education*, 1958, *33*, 491–500.

Karren, R. Alternative selection devices: The work sample and the interview. Paper presented at National Council of Measurement in Education Convention, Boston, Massachusetts, April 1980.

Katz, M. R. (review). In O. K. Buros, *Seventh mental measurements yearbook*, Volume II. Highland Park, New Jersey: The Gryphon Press, 1972, 1462–1465.

Keck, J. W.; Arnold, L.; Willoughby, L.; and Calkins, V. Efficacy of cognitive/noncognitive measures in predicting resident-physician performance. *Journal of Medical Education*, 1979, *54*, 759–765.

Kegel-Flom, P. Predicting supervisor, peer, and self-ratings of intern performance. *Journal of Medical Education*, 1975, *50*(7–12), 812–815.

Kelly, E. L. Theory and techniques of assessment. *Annual Review of Psychology*, 1954, *5*, 286–290.

Kelly, E. L. A critique of the interview. *Journal of Medical Education*, 1957(a), *32*(10:2), 78–84.

Kelly, E. L. Multiple criteria of medical education and their

implications for selection. *Journal of Medical Education*, 1957(b), *32*(10:2), 185–196.

Kelly, E. L., and Fiske, D. *The prediction of performance in clinical psychology*. Ann Arbor, Michigan: University of Michigan Press, 1951.

Keochakian, S. V. *Nonintellectual factors in admissions: Access policy and procedures and the law in U.S. higher education*. New York: International Council for Educational Development, Number 6, 1978, 95–111.

Knapp, J., and Sharon, A. *A compendium of assessment techniques*. Cooperative Assessment of Experiential Learning. Princeton: Educational Testing Service, 1975.

Korman, M. A. A factorial study of judgmental space. *Psychological Reports*, 1962, *10*, 739–746.

Korman, M.; Stubblefield, R. L.; and Martin, L. W. Patterns of success in medical school and their correlates. *Journal of Medical Education*, 1968, *43*, 405–411.

Krug, R. E. An analysis of eighteen Peace Corps projects. *American Institutes for Research*, October 1962, 16–20.

Kryger, R. S., and Shikiar, R. Sexual discrimination in the use of letters of recommendations: A case of reverse discrimination. *Journal of Applied Psychology*, 1978, *63*(3), 309–314.

LaGaipa, J. J. *Development of officer candidate biographical information blank: Vol. I: Validation of some experimental inventories*. Washington, D.C.: Bureau of Naval Personnel Technical Bulletin, 60-9, June 1960.

Lavin, D. E. *The prediction of academic performance: A theoretical analysis and review of research*. New York: Russell Sage Foundation, 1965.

Leape, L. L.; Palubinskas, A. L.; Steindler, J.; Wild, B.; and Dalrymple, W. Peer evaluation of applicants to medical school. *Journal of Medical Education*, 1976, *51*, 586–588.

Lenning, O. T.; Munday, L. A.; Johnson, O. B.; Vander Well, A. R.; and Brue, E. J. *The many faces of college success and their nonintellective correlates: The published literature through the decade of the sixties*. American College Testing Program, Monograph No. 15, 1974.

Levitt, M. That bouillabaisse: Medical school admissions. In H. S. Sacks and associates, *Hurdles: The admissions dilemma in American higher education*. New York: Atheneum, 1978, 239–266.

Lewin, A. Y., and Zwany, A. Peer nominations: A model, literature critique and a paradigm for research. *Personnel Psychology*, 1976, *29*, 423–447.

Lewis, L. S. On the genesis of gray-flanneled puritans. AAUP *Bulletin*, 1972, *58*, 21–29.

Lief, V. F.; Lief, H. I.; and Young, K. M. Academic success: Intelligence and personality. *Journal of Medical Education*, 1965, *40*, 114–124.

Lindsay, C. A., and Althouse, R. Comparative validities of the Strong Vocational Interest Blank academic achievement scale and the College Student Questionnaire motivation for grades scale. *Educational and Psychological Measurement*, 1969, *29*, 489–493.

Linn, R. L., and Winograd, P. A. New York University admissions interview study. Report No. LSAC-69-2. In Law School Admission Council, *Reports of* LSAC *Sponsored Research: Volume 1, 1949–1969*. Princeton: Law School Admission Council, 1969.

Litton-Hawes, E.; MacLean, I. C.; and Hines, M. H. An analysis of the communication process in the medical admissions interview. *Journal of Medical Education*, 1976, *51*, 332–334.

Livingston, S. A. *Nonverbal communication tests as predictors of success in psychology and counseling*. Report (COPA-80-01). Center for Occupational and Professional Assessment. Princeton: Educational Testing Service, June 1980.

Lohnes, P. R. (review). In O. K. Buros, *Seventh mental measurements yearbook*, Volume II. Highland Park, New Jersey: The Gryphon Press, 1972, 1423–1424.

Lunneborg, P. W., and Lillie, C. Sexism in graduate admissions: The recommendation. *American Psychologist*, 1973, *28*(2), 187–189.

Lunneborg, P. W., and Lunneborg, C. E. The differential prediction of college grades from biographic information. *Educational and Psychological Measurement*, 1966, *26*, 917–925.

Malloy, J. P. The prediction of college achievement with the life experience inventory. *Educational and Psychological Measurement*, 1955, *15*, 170–180.

Malloy, J. P., and Ivanoff, J. M. Further use of the life experience inventory in predicting college achievement. *Journal of Educational Research*, 1964, *57*(10), 522–525.

Mayfield, E. C. The selection interview: A re-evaluation of published research. *Personnel Psychology*, 1964, *17*, 239–260.

Mayfield, E. C. Value of peer nominations in predicting life insurance sales performance. *Journal of Applied Psychology*, 1972, *56*(4), 319–323.

McCausland, D. F., and Stewart, N. E. Academic aptitude, study skills, and attitudes and college GPA. *Journal of Educational Research*, 1974, *67*(8), 354–357.

McClelland, D. C. *The achieving society*. New York: The Free Press, 1961.

McCormick, E. J., and Ilgen, D. *Industrial psychology*, 7th ed. Englewood Cliffs, New Jersey: Prentice-Hall, Inc., 1980.

McGuire, C. H., and Babbott, D. Simulation technique in the measurement of problem-solving skills. *Journal of Educational Measurement*, 1967, *4*(1), 1–10.

Medland, F. F.; Yates, L. G.; and Downey, R. G. Associate ratings of senior officer potential. *Research Problem Review*,

74-2, U.S. Army, Research Institute for the Behavioral and Social Sciences, 1974.

Mehrabian, A. Undergraduate ability factors in relationship to graduate performance. *Educational and Psychological Measurement*, 1969, *29*(2), 409–419.

Mendel, R. W., and Tabb, W. G. Problems in admissions in U.S. dental schools. *College and University*, 1977, *52*(3), 299–316.

Messick, S. Personality measurement and college performance. In S. J. Messick and D. N. Jackson (eds.), *Problems in Human Assessment*. New York: McGraw-Hill, 1967, 834–845.

Messick, S. Potential uses of noncognitive measurement in education. *Journal of Educational Psychology*, 1979, *71*(3), 281–292.

Milstein, R. M.; Wilkinson, L.; Burrow, G. N.; and Kessen, W. Admissions decisions and performance during medical school. *Journal of Medical Education*, 1981, *56*(2), 77–82.

Mitter, W. *The use of tests and interviews for admission to higher education*. Windsor, England: NFER Publishing Company, 1979.

Moore, L. F., and Lee, A. J. Comparability of interview, group, and individual interview ratings. *Journal of Applied Psychology*, 1974, *59*(2), 163–167.

Morrison, R. F.; Owens, W. A.; Glennon, J. R.; and Albright, L. E. Factored life history antecedents of industrial research performance. *Journal of Applied Psychology*, 1962, *46*, 281–284.

Morse, P. K., and Moebes, J. D. Effect of interview on admissions decisions. *Journal of Dental Education*, 1973, *37*(5), 26.

Mosel, J. N., and Goheen, H. W. The validity of the employment recommendation questionnaire in personnel selection: I. Skilled trades. *Personnel Psychology*, 1958, *2*, 481–490.

Mosel, J. N., and Goheen, H. W. The employment recommendation questionnaire: III. Validity of different types of references. *Personnel Psychology*, 1959, *12*, 469–477.

Moses, J. L., and Byham, W. C. *Applying the assessment center method*. Elmsford, New York: Pergamon Press, 1977.

Mount, M. K.; Muchinsky, P. M.; and Hanser, L. M. The predictive validity of a work sample: A laboratory study. *Personnel Psychology*, 1977, *30*, 637–645.

Murden, R.; Galloway, G. M.; Reid, J. C.; and Colwill, J. M. Academic and personal predictors of clinical success in medical school. *Journal of Medical Education*, 1978, *53*, 711–719.

Myers, R. C. Biographical factors and academic achievement: An experimental investigation. *Educational and Psychological Measurement*, 1952, *12*, 414–426.

Myers, R. C., and Schultz, D. G. Predicting academic achievement with a new attitude-interest questionnaire: I. *Educational and Psychological Measurement*, 1950, *10*, 654–663.

Nevis, E. C. The effectiveness of the Worthington personal history technique in assessing air force officers for command and staff leadership. *Bibliography of Publications and Abstracts of Dissertations*. Cleveland, Ohio: Case Western Reserve University, 1954.

Nichols, R. C. Nonintellective predictors of achievement in college. *Educational and Psychological Measurement*, 1966, *26*, 899–915.

Nichols, R. C., and Holland, J. L. Prediction of the first year college performance of high aptitude students. *Psychological Monographs*, 1963, *77*(7), 1–29.

Nicholson, E. *Success and admission criteria for potentially successful risks*. Providence, Rhode Island: Brown University, 1970.

Norris, L., and Katz, M. R. *The measurement of academic interests: Part II. The predictive validities of academic interest measures*. (RB-70-67), 1970.

O'Sullivan, M., and Guilford, J. P. Six factors of behavioral cognition: Understanding other people. *Journal of Educational Measurement*, 1975, *12*, 255–271.

Owens, W. A. Toward one discipline of scientific psychology. *American Psychologist*, 1968, *23*, 782–785.

Owens, W. A. A quasi-actuarial basis for individual assessment. *American Psychologist*, 1971, *26*(2), 992–999.

Owens, W. A. Background data. In M. D. Dunnette (ed.), *Handbook of industrial and organizational psychology*. Chicago: Rand McNally, 1976.

Owens, W. A., and Schoenfeldt, L. F. Toward a classification of persons (monograph). *Journal of Applied Psychology*, 1979, *65*(5), 569–607.

Pace, L. A., and Schoenfeldt, L. F. Legal concerns in the use of weighted applications. *Personnel Psychology*, 1977, *30*(2), 159–166.

Palacios, M. H.; Newberry, L. A.; and Boozin, R. R. Predictive validity of the interview. *Journal of Applied Psychology*, 1966, *50*(1), 67–72.

Payne, D. A.; Rapley, F. E.; and Wells, R. A. Application of a biographical data inventory to estimate college academic achievement. *Measurement and Evaluation in Guidance*, 1973, *6*(3), 152–156.

Petry, J. R., and Craft, P. A. Investigation of instruments to predict academic performance of high-risk college students. *Journal of Educational Research*, 1976, *70*(1), 21–25.

Prediger, D. J. Prediction of persistence in college. *Journal of Counseling Psychology*, 1965, *12*, 62–67.

Price, P. B.; Lewis, E. G.; Loughmiller, G. C.; Nelson, D. E.; Murray, S. L.; and Taylor, C. W. Attributes of a good practicing physician. *Journal of Medical Education*, 1971, *46*, 229–237.

Quann, J. C. *Admissions, academic records, and registrar services.* San Francisco: Jossey-Bass Publishers, 1979.

Reich, S. The relationship of personality characteristics and vocational interests to academic achievement in a school of law. *Dissertation Abstracts International, 33,* (1-2), 449-B, Fordham University, 1972.

Reilly, R. R. The reliability and validity of a standard law school recommendation form. Report No. LSAC-71-5. In Law School Admission Council, *Reports of* LSAC *Sponsored Research:* Volume II (1970-74). Princeton: Law School Admission Council, 1976, 149-164.

Reilly, R. R., and Manese, W. R. The validation of a mini-course for telephone company switching technicians. *Personnel Psychology,* 1979, *32,* 83-90.

Ricciuti, H. Ratings of leadership potential at the U.S. Naval Academy and subsequent officer performance. *Journal of Applied Psychology,* 1955, *39,* 194-199.

Richards, J. M., Jr.; Culkins, E. V.; McCanse, A.; and Burgess, M. M. Predicting performance in a combined undergraduate and medical education program. *Educational and Psychological Measurement,* 1974, *34,* 923-931.

Richards, J. M.; Holland, J. L.; and Lutz, S. W. The prediction of student accomplishment in college. *Journal of Educational Psychology,* 1967, *58*(6), 343-355.

Richards, J. M., Jr., and Lutz, S. W. Predicting student accomplishment in college from the ACT assessment. *Journal of Educational Measurement,* 1968, *5*(1), 17-29.

Richards, J. M., Jr., and Taylor, C. W. Predicting academic achievement in a college of medicine from grades, test scores, interviews, and ratings. *Educational and Psychological Measurement,* 1961; *21*(4), 987-994.

Richards, R. L., and Casey, M. B. *Creativity and academic motivation variables as predictors of achievement in a two-year college for educationally marginal students.* Paper presented to annual meeting of the American Educational Research Association, Washington, D.C., March 30-April 3, 1975.

Rim, Y. How reliable are letters of recommendation? *Journal of Higher Education,* 1976, *47*(4), 437-445.

Roadman, H. E. An industrial use of peer ratings. *Journal of Applied Psychology,* 1964, *48,* 211-214.

Rowell, J. A., and Renner, V. J. Personality, mode of assessment, and student achievement. *British Journal of Educational Psychology,* 1975, *45,* 232-236.

Samph, T., and Best, A. M., III. An experimental videotape examination of medical student interpersonal skills. Paper presented at the annual meeting of the American Educational Research Association, San Francisco, April 1979.

Schaefer, C. E. The prediction of creative achievement from a biographical inventory. *Educational and Psychological Measurement,* 1969, *29,* 431-437.

Schmidt, F.; Greenthal, A. L.; Hunter, J. E.; Berner, J. B.; and Seaton, F. W. Job sample vs. paper-and-pencil trades and technical tests: Adverse impact and examinee attitudes. *Personnel Psychology,* 1977, *30,* 187-198.

Schoenfeldt, L. F. Utilization of manpower: Development and evaluation of an assessment-classification model for matching individuals with jobs. *Journal of Applied Psychology,* 1974, *59*(5), 583-595.

Schultz, D. G., and Green, B. F., Jr. Predicting academic achievement with a new attitude-interest questionnaire—II. *Educational and Psychological Measurement,* 1953, *13,* 54-64.

Scott, O., and Castles, J. Aptitude motivation and life history as predictors of the nonacademic accomplishments of high school seniors. *Measurement and Evaluation in Guidance,* 1975, *7*(1), 37-42.

Segal, S. M.; Busse, T. V.; and Mansfield, R. S. The relationship of scientific creativity in the biological sciences to predoctoral accomplishments and experiences. *American Educational Research Journal,* 1980, *17*(4), 491-502.

Shaffer, D. R.; Mays, P. V.; and Etheridge, K. Who shall be hired: A biasing effect of the Buckley Amendment on employment practices? *Journal of Applied Psychology,* 1976, *61*(5), 571-575.

Skager, R. W.; Schultz, C. B.; and Klein, S. P. Quality and quantity of accomplishments as measures of creativity. *Journal of Educational Psychology,* 1965, *56,* 31-39.

Smith, G. M. Usefulness of peer ratings of personality in educational research. *Educational and Psychological Measurement,* 1967, *27*(4), 967-984.

Smith, L. Significant differences between high-ability achieving and nonachieving college freshmen as revealed by interview data. *The Journal of Educational Research,* 1965, *59*(1), 10-12.

Smith, W. J.; Albright, L. E.; Glennon, J. R.; and Owens, W. A. The prediction of research competence and creativity from personal history. *Journal of Applied Psychology,* 1961, *45*(1), 59-62.

Splitstone, A. The admissions interview. *College and University,* 1974, *49*(4), 337-342.

Stein, M. I. *Personality measures in admissions.* New York: College Entrance Examination Board, 1963.

Stricker, L. J. *Interpersonal competence instrument: Development and preliminary findings.* Princeton: Graduate Management Admissions Council, January 1980.

Stroup, A. L. The prediction of academic performance from personality and aptitude variables. *The Journal of Experimental Education,* 1970, *38*(3), 83-86.

Taylor, C. W., and Ellison, R. L. Biographical predictors of scientific performance: Criteria of scientific performance

and creativity can be predicted from biographical information. *Science*, March 1967, 1075–1080.

Taylor, H. C., and Russell, J. T. The relationship of validity coefficients to the practical effectiveness of tests in selection: Discussion and tables. *Journal of Applied Psychology*, 1939, *23*, 565–578.

Thomas, J. A. Heavy traffic on the purple brick road: The route to law school. In H. S. Sacks and associates, *Hurdles: The admissions dilemma in American higher education*. New York: Atheneum, 1978, 212–238.

Tucker, D. H., and Rowe, P. M. Relationship between expectancy, causal attributions, and final hiring decisions in the employment interview. *Journal of Applied Psychology*, 1979, *64*(1), 27–34.

Tullar, W. L., and Barrett, G. V. The future autobiography as a predictor of sales success. *Journal of Applied Psychology*, 1976, *61*(3), 371–373.

Tupes, E. C. Relationships between behavior trait ratings by peers and later officer performance of USAF OCS graduates. Personnel Laboratories, Air Force Personnel and Training Research Center, Air Research and Development command, Lackland Air Force Base, Texas. Research Report AFPTRC-TN-57-125, ASTIA Document No. 134257, October 1957.

Ulrich, L., and Trumbo, D. The selection interview since 1949. *Psychological Bulletin*, 1965, *63*(2), 100–116.

Wagner, R. The employment interview: A critical summary. *Personnel Psychology*, 1949, *2*, 17–46.

Wallach, M. A. The psychology of talent and graduate education. In S. Messick (ed.), *Individuality in learning: Implications of cognitive styles and creativity for human development*. San Francisco: Jossey-Bass, 1976.

Wallach, M. A., and Wing, C. W. *The talented student: A validation of the creativity-intelligence distinction*. New York: Holt, Rinehart, and Winston, 1969.

Walsh, W. B. Validity of self-report. *Journal of Counseling Psychology*, 1967, *14*(1), 18–23.

Walsh, W. B. Validity of self-report: Another look. *Journal of Counseling Psychology*, 1968, *15*(2), 180–186.

Walsh, W. B. Self-report under socially desirable and distorted conditions. *Journal of Counseling Psychology*, 1969, *16*(6), 569–574.

Walsh, W. B. (review). In O. K. Buros, *Seventh mental measurements yearbook*, Volume II. Highland Park, New Jersey: The Gryphon Press, 1972, 1429–1431.

Walsh, W. B., and Hanle, N. A. Consistent occupational preferences, vocational maturity, and academic achievement. *Journal of Vocational Behavior*, 1975, *7*, 89–97.

Ward, L. B. The interview as an assessment technique. In *College Admissions*. New York: College Entrance Examination Board, 1955, 62–71.

Ward, W. C.; Frederiksen, N.; and Carlson, S. B. Construct validity of free-response and machine scorable forms of a test. *Journal of Educational Measurement*, 1980, *17*(1), 11–29.

Waters, L. K., and Waters, C. W. Peer nominations as predictors of short-term sales performance. *Journal of Applied Psychology*, 1970, *54*(1), 42–44.

Watson, C. G. California Psychological Inventory as a predictor of academic achievement in normal and maladjusted college males. *Journal of Educational Research*, 1967, *61*(1), 10–13.

Watson, C. G. Cross-validation of certain background variables as predictors of academic achievement. *Journal of Educational Research*, 1965, *59*(4), 147–148.

Webb, S. C. The comparative validity of two biographical inventory keys. *Journal of Applied Psychology*, 1960, *44*(3), 177–183.

Webster, E. C. *Decision-making in the employment interview*. Montreal: Eagle Publishing Co., 1967, 61–71.

Weinstein, E. L.; Brown, I.; and Wahlstrom, M. W. A systems view of admissions procedures. *College and University*, 1979, *54*(2), 124–138.

Weitz, H., and Wilkerson, H. J. The relationship between certain nonintellective factors and academic success in college. *Journal of Counseling and Psychology*, 1957, *4*(1), 54–60.

Wells, P. Applying to college: Bulldog bibs and potency myths. In H. S. Sacks and associates, *Hurdles: The admissions dilemma in American higher education*. New York: Atheneum, 1978, 48–75.

Whitaker, W.; Kelly, E. L.; and Uhr, L. The Michigan project on the prediction of professional competence in medicine. *Journal of Medical Education*, 1957, *32*(10:2), 187–188.

Whitla, D. K. Evaluation of decision making: A study of college admissions. In D. K. Whitla (ed.), *Handbook of measurement and assessment in behavioral sciences*. Reading, Massachusetts: Addison-Wesley, 1968.

Whitney, D. R. Predicting from expressed vocational choice: A review. *Personnel and Guidance Journal*, 1969, *48*, 279–286.

Wickenden, J. W., Jr. Memorandum to all Princeton alumni on the admissions process. *Princeton Alumni Weekly*, October 22, 1979.

Williams, M. Faculty and students as admissions raters: A comparison of the validity of their predictions. *College and University*, 1978, *53*(2), 172–182.

Willingham, W. W. The application blank as a predictive instrument. *College and University*, 1965, *40*, 271–278.

Willingham, W. W. Predicting success in graduate education. *Science*, 1974, *183*, 273–278.

Wilson, K. M. The utility of standard composite for academic performance in several liberal arts colleges. *Research in Higher Education*, 1976, *5*, 193–213.

Wing, C. W., Jr., and Wallach, M. A. *College admissions and the psychology of talent*. New York: Holt, Rinehart, and Winston, 1971.

Witkin, H. A.; Moore, C. A.; Goodenough, D. R.; and Cox, P. W. Field-dependent and field-independent cognitive styles and their educational implications. *Review of Educational Research*, 1977a, *47*, 1–64.

Witkin, H. A.; Moore, C. A.; Oltman, P. K.; Goodenough, D. R.; Friedman, F.; Owen, D. R.; and Raskin, E. Role of the field-dependent and field-independent cognitive styles in academic evolution: A longitudinal study. *Journal of Educational Psychology*, 1977b, *69*(3), 197–211.

Wollowick, H. B., and McNamara, W. J. Relationship of the components of an assessment center to management success. *Journal of Applied Psychology*, 1969, *53*(5), 348–352.

Wright, O. R., Jr. Summary of research on the selection interview since 1964. *Personnel Psychology*, 1969, *22*, 391–413.

Annotated bibliography

Annotations were made for studies either reporting evidence or providing rationales related to any of the five procedures emphasized in this review: biodata, recommendations, interviews, interest measures, and personality measures. Since some citations may have dealt with more than one of the procedures, multiple annotations occur for some of them. The annotations are presented in order of procedures reviewed, as listed above.

Biodata

Aiken (1964)

The author devised a 76-item, multiple-choice biographical inventory for predicting academic performance and attrition, and this inventory was administered to 1,006 incoming freshmen at the Women's College of the University of North Carolina. Two random samples of 100 each were drawn for developing scoring keys for the university (Group 1) and for cross-validation (Group 2). The biographical inventory correlated .60 with first-semester GPA in Group 1 and .57 in Group 2. Multiple correlations were increased from .54 to .68 in Group 1 and from .58 to .70 in Group 2 using the inventory (SAT-V, SAT-M, and high school class rank were the other predictors). Attrition was studied by comparing 50 freshman women who withdrew from college before or at midsemester with 50 randomly selected nonwithdrawals. It was noted that the items that predicted attrition had only a small overlap with those that predicted academic performance.

Anastasi (1976)

This book relates the biographical inventory to the interview by noting that both are important as sources of life-history data, but that biographical inventories are at times more feasible than interviews, and that they are often used in conjunction with the interview for preliminary data gathering. The author observes that while both biographical inventories and personality inventories are self-report instruments, the former is more objective and the information obtained is readily verifiable. For maximum predictive effectiveness, the author recommends that inventory items be selected and weighted by criterion keying and then be cross-validated against the criterion in a new sample. When such procedures are followed, it is concluded that "biographical inventories have proved to be consistently good predictors of performance in a wide variety of contexts" (page 615). Examples given are predicting job turnover of bank clerks, amount of insurance sold by life insurance salesmen, productivity of research scientists, and artistic creativity in high school students.

Anastasi, Meade, and Schneiders (1960)

This monograph describes the development and validation of a scoring key for a biographical inventory intended for use as a predictor of college success. The criterion of college success emphasized nonintellective factors, but academic achievement was also recognized.

Entering freshman classes at Fordham University in 1958 and 1959 were studied. Based on their performance during the first three years of college, these students were classified into three criterion groups: positive, average, and negative. These classifications were made by a committee of three judges on the basis of information assembled from nine different sources. It was concluded that the biographical inventory was a better predictor of the criterion developed than the usual college aptitude tests.

Anastasi and Schaefer (1969)

A biographical inventory was developed for the identification of creativity in art and writing. Teacher nominations were used to identify creative students among 400 high school girls. Matched control groups were identified and scored on the biographical inventory, keyed either for creative art or creative writing, and scores were computed for each of the four groups. Mean scores for the creative groups, both in art and in writing, were significantly higher than scores for the respective control groups. A dichotomous criterion of creative vs. control was also used, and point-biserial correlations between the biographical inventory scores and the dichotomous criterion were .34 for the art group and .55 for the writing group. The biographical inventory consisted of 166 items with sections on physical characteristics, family history, educational history, leisure-time activities, and miscellaneous.

Baird (1969a)

A sample of 12,432 college students in 31 institutions was administered the American College Survey during the spring of 1964. Of these, 5,129 students at 29 institutions participated in a follow-up study. Records of the college accomplishments of these students were obtained in the spring of 1965 at the end of their sophomore year. High school achievement and college achievement of a comparable nature were then related, and it was shown that a number of primarily nonacademic college achievements could be predicted from high school achievement of similar types. Music achievement and art achievement in college were predicted quite well (R = .57 for men and R = .47 for women). Slightly less accurate predictions were made for leadership achievement in college (R = .37 for men and R = .42 for women).

Baird (1969d)

A checklist of extracurricular accomplishments was developed, and scores on leadership, music, drama and speech, art, writing, and science were developed from it. These scores were related both to grades and to nonclassroom achievement in two-year colleges using samples of 1,628 men and 1,079 women. The study confirmed earlier findings that such scores are useful in the prediction of nonclassroom achievement. For example, college art achievement was predicted for women (multiple R = .61) and for men (multiple R = .46). Worst predictions were for college science achievement for women (multiple R = .17), even though the predictive relationship for men was somewhat better (multiple R = .34).

Baird (1976)

This monograph represents a good review of the uses and validity of self-reported information, such as that provided on application forms, biographical inventories, and other brief self-report assessments. The report argues that, in contrast to traditional tests, relatively little attention has been given to self-reported data even though biographical information has consistently been found to predict later achievement. The monograph has sections covering the prediction of academic performance, the prediction of creative and high-level accomplishment, and the prediction of vocational choice and success. It is suggested that the particular advantages of brief self-report assessment include "the breadth of its coverage, its utility in a variety of decision-making situations, its potential use as a learning experience, its ability to screen large numbers of people inexpensively and quickly, and the ease with which it is obtained" (page 72).

Baird (1979)

The first two stages are described of a four-stage project designed to develop procedures for the assessment of the accomplishments of applicants to graduate schools. The first stage was the development of trial inventories based on previous attempts. In the second stage of the

project an inventory was developed and reviewed by diverse representatives concerned with graduate admissions. The third stage of the project will be a test of the instrument in a sample of graduate school departments, and the fourth stage will be a broader tryout including a long-term predictive validity evaluation. The report includes a copy of the Inventory of Activities and Accomplishments, the instrument being examined in the third stage of the project.

Brown and Dubois (1964)

Two samples of male students who were freshmen at the University of Iowa during the 1961–62 academic year were administered a number of instruments including a biographical inventory. Both predictor and criterion (cumulative GPA for the first two quarters) data were obtained for 76 science and humanities students and 125 engineering students. The biographical inventory included questions on the number of hours studied per week, hours spent thinking about classwork between classes, hours of free time between classes spent studying, student's lowest acceptable GPA, and parent's lowest acceptable GPA. The highest predictive correlations were obtained with the last two questions on student and parent acceptable GPAs. Hours studied predicted cumulative GPA for engineering students significantly.

England (1961)

This book offers a systematic, but dated, treatment of the development and use of weighted application blanks in employment. Special emphasis is given to the procedural steps for deriving item weights. Important considerations for use of weighted application blanks are also covered. Evidence of the validity of weighted application blanks for predicting employee tenure, salary increases, and success ratings is given and studies cited. The weighted application blank is viewed as especially useful for situations in which the number of applicants is considerably greater than the number of places available or in which it is expensive to bring applicants to some central facility for testing or interviewing. Evidence presented strongly suggests that a well-developed application blank can lead to better decisions than are made on the basis of short interviews. It is argued, however, that the weighted application blank is not a substitute for other selection tools but an additional tool which should increase the accuracy of decisions when used with other techniques.

Freeberg (1976)

This paper reviews past and prospective uses of the biographical information blank as a predictor of academic achievement, creative achievement, persistence in school, and vocational and curriculum choice. The author observes that the use of biographical information as a predictor is rooted in the industrial setting but that some interest is evolving in education as a result of more broadly defined goals. Whereas educational prediction has traditionally focused narrowly on academic achievement as an outcome, newer concerns are the prediction of nonclassroom achievement, long-range creative achievement, and persistence in college. The book reports evidence that biographical information predicts academic achievement but emphasizes that such information requires a broader perspective in order to fully utilize its potential. The review notes practical advantages of biographical information such as low cost and its relatively innocuous nature, especially when obtained as an extension of the application form.

Harrington (1969)

The purpose of this study was to determine the extent to which biographical data provide an effective means of predicting freshman academic performance. Subjects were 746 men and 780 women freshmen who entered Ohio University in the fall of 1966. All subjects were administered a 300-item biographical inventory that yielded two scales: a creativity scale and a management potential scale. ACT test scores, high school rank, and first-semester freshman GPA were other measures used. The ACT Math score yielded the highest correlations with GPA for both men (.43) and women (.58). The ACT composite correlated slightly less with GPA (.32 for men; .42 for women). High school rank predicted GPA somewhat better than the ACT composite (.43 for men; .50 for women). The creativity scale was somewhat less predictive of GPA (.27 for men; .30 for

women), and the management potential scale even less so (.21 for men; .17 for women). The overall multiple correlations were .54 for men and .66 for women. The results indicated that the biographical inventory could provide useful information for the prediction of college performance.

Henry (1966)

This is a report of a conference held in 1965 to explore past and future research on the biographical inventory. It is reported that, with few exceptions, the biographical inventory has been found to be the best single predictor of future behavior when that behavior is of a complex nature, such as that required of a business executive. It is noted that the biographical inventory is conceived of as an extended structured interview on paper. The conclusions of the conference were that there was a need for better understanding of biodata and their uses and for a clearinghouse of information related to biodata.

Hilton and Myers (1967)

A 169-item Background and Experience Questionnaire (BEQ) was completed by all twelfth-grade students in 17 public school systems in the spring of 1963 (N = 1,206). The objective of the investigation was to determine if a biographical inventory could contribute uniquely to the prediction of twelfth-grade academic standing over and beyond that provided by eleventh-grade test performance. Three tests, the School and College Ability Test (SCAT), the Sequential Test of Educational Progress (STEP), and the Test of General Information (TGI) were used. It was concluded that, while the BEQ correlated significantly with twelfth-grade rank, the increase in prediction provided beyond that possible with aptitude test scores was slight.

Korman (1968)

Sixty-eight students of the 1962 class at the University of Texas Southwestern Medical School (Dallas) were followed through medical school, and admissions variables were related to factors based on academic performance, clinical performance, peer ratings, and internship performance. Medical school performance through the senior year was included. Biographical data were used to predict five performance factors: general achievement (including clinical performance), internship success, peer esteem, scientist potential, and humanism. Significant correlations were obtained between biographical items and all of these long-term performance criteria, but different items were related to different criteria. For example, internship success was predicted by "frequency of teen dating" (r = .31), "many friends" (r = .27), and "parental interest" (.28).

Lunneborg and Lunneborg (1966)

The goal of this study was to identify useful biographical items from data already at hand in admissions or in application forms. Subjects were 520 freshmen at the University of Washington in the fall of 1965. Predictors consisted of high school GPAs in different subject areas, 12 test scores, 71 items from the application form, and 62 items from a Survey of College Plans. Criteria were first-quarter overall GPA, English GPA, math GPA, foreign language GPA, physical science GPA, change in major at the end of the first quarter, and withdrawal from school. A sequential predictor selection was used to identify the most important predictors for each criterion. Of the first six predictors for each of the seven criteria, only one high school grade average and only five tests scores were selected. In contrast, 22 of 32 biographic variables were selected as best predictors. Withdrawal from college was predicted with five biographical predictors (r = .35). Eight biographical variables predicted change of major with an r = .35.

Malloy (1955)

This investigation was intended to identify biographical experiences and attitudes associated with first-semester academic achievement and to determine the effectiveness of a biographical inventory when used in conjunction with preregistration test scores. Approximately 1,500 Life Experience Inventories (LEIs) were distributed to freshman classes at the University of Nebraska during the second week of the 1952 fall semester. A total of 875 of the completed LEIs were used in the study, and these were divided into an initial group of 434 and a cross-validation group of 441. The LEI correlated .43 with first-semester grades for men and .45 for women. The multiple correlation of all predictors was .56 for men and .59 for women. It was concluded that the LEI significantly increased the predic-

tion of college performance over that possible with the placement tests then in use. The LEI was described as an inventory sampling from school experiences, self-appraisal, family relationships, and choice and type of friends.

Malloy and Ivanoff (1964)

In this study, the Life Experience Inventory (LEI) was subjected to further evaluation in response to a criticism (Fishman and Pasanella, 1960) that an earlier validation study (Malloy, 1955) had failed to include high school achievement as a predictor. Subjects were 60 men and 80 women enrolled at Marquette University. The dependent variable was accumulated GPA at the end of the sophomore year. Multiple correlations obtained between the five independent variables and the criterion were .69 for men and .70 for women. The LEI score was the best single predictor for men ($r = .56$). It was concluded that the LEI explained unique criterion variance over and above that accounted for by tests and the high school record.

Myers (1952)

The purpose of this investigation was to determine whether biographical information was useful in predicting performance at an Eastern women's liberal arts college. Both the 1951 and the 1952 entering classes were studied. A total of 355 freshmen in the 1951 class and 346 freshmen in the 1952 class were administered biographical inventories, and seven items out of the inventory were selected because they were shown to be related to performance after academic aptitude was partialed out. The seven biographical items selected were based on religion, extent of participation in outside activities, county of residence, birthplace of parents, birth order, and number of high school offices held. A biographical score based on these seven questions correlated .26 with freshmen GPA in 1951 and .29 in 1952. Partial predictive correlations, with aptitude scores and high school performance removed, were .25 and .19 for 1951 and 1952, respectively. It was concluded that the biographical score added a small but significant contribution to the prediction of freshman academic performance.

Myers and Schultz (1950)

This paper reports on efforts to develop an instrument for measuring nonintellectual factors associated with subsequent academic achievement in college. Recognizing that previous attempts to use attitude, interest, motivation, and personality inventories had been discouraging, the authors decided to construct a questionnaire covering a wide range of factors. The original questionnaire of 328 items was first administered to freshmen entering an Eastern women's liberal arts college in 1947 (the class of 1951). The questionnaire was shortened to 145 items after initial experimentation. Scoring keys were then developed from an over/under-achiever residual index. Modest correlations were obtained between three scales developed and freshman GPA. However, no significant addition to a multiple correlation based on SAT scores and the high school record was achieved with the questionnaire.

Nichols and Holland (1963)

As a part of a larger study, National Merit Finalists who graduated in 1960 were asked a number of questions about their background. Responses were then correlated with academic and nonacademic achievements during the first year of college. Specifically, students were asked to provide information of the following types: ratings of artistic achievement in high school, ratings of scientific achievement in high school, number of elective offices held in high school, and breadth of creative activity assessed by a 32-item checklist. These predictors were correlated with both specific college achievements and a composite of nonacademic college achievements. It was concluded that high school achievements—in contrast to a number of interest, attitude, personality, and other measures—were the best predictors of nonacademic college achievements.

Nicholson (1970)

Two types of biographical data were used to predict short- and long-range outcomes of students who had attended Brown University. An environmental background index (EI) was developed by summing data

typically found in personal information forms required of candidates for admission, related items from the College Student Questionnaire (CSQ), and other items of admissions and guidance information. A second type of biographical data was developed from a survey of activities and accomplishments. Two variables of activities were devised: LAS (leadership, activities, and social participation) and LOMAS (literary, oratorical, musical, artistic, and scientific participation). Test-retest reliabilities of EI, LAS, and LOMAS were found to equal or exceed .80. EI was found to discriminate significantly between nonveterans of 1950–52 classes rated as successful or nonsuccessful by peers 15 years later, those who obtained advanced degrees vs. those who obtained no advanced degree, and graduates vs. nongraduates. For veterans, EI only discriminated significantly between graduates and nongraduates. For 1966–68 classes, EI discriminated significantly between graduates and nongraduates and between advanced degree graduates. Results with the activities indexes were less consistent but indicated that culturally related activities discriminated significantly between long-term success and nonsuccess groups, but athletically related activities did not.

Owens (1968)

This paper presents a theoretical rationale for biographical data as a link between psychometric and experimental psychology. Owens argues that the biographical inventory can provide information needed for better understanding of correlations found among variables used in predicting and measuring performance and that generalizations about experimental effects can be tested for different groups identified by the biographical inventory.

Owens (1971)

This paper describes a model and presents evaluation data in support of the use of biodata to classify individuals and to predict behavior. Although the model described has been tested on college students, other studies are cited of similar uses of biodata in business and industrial settings to identify salesmen and executives. The particular value of biodata as an efficient method of identification is emphasized.

Owens and Schoenfeldt (1979)

Results of a long-term investigation of life history, or scored biographical data, at the University of Georgia are reported. An extensive study of 2,000 biodata items led to the development of a 118-item inventory, the University of Georgia Biographical Questionnaire (BQ). The BQ was administered to virtually all freshmen entering the university for the years 1970, 1971, 1972, and 1973. The same 118-item BQ was administered also to all freshmen at a technically oriented college in the Northeast, and the factor structure was closely duplicated. The BQ was then used in combination with reference measures such as high school GPA and SAT scores and college behavior to identify 23 male and 15 female subgroups through factor analysis. Male groups included these descriptive types: "jocks," "farm boys," "eggheads," leaders, business-oriented, "fraternity Joes," and authoritarian dropouts. Female groups included: well-adjusted achievers, scholarly bookworms, cognitively simple emotional dropouts, and mature liberal humanists. It was noted that 50 percent of the members of one subgroup drop out of college during their first year as compared to only 7 percent of another subgroup, but these subgroups were not identified specifically.

Payne, Rapley, and Wells (1973)

A 389-item biodata questionnaire developed at the University of Georgia (see Owens and Schoenfeldt, 1979) was used to predict freshman GPA for males and females separately. The most discriminating 275 items of the questionnaire were factored into 13 male and 15 female factors. One factor, academic achievement, correlated highest of all factors with freshman GPA for both males and females. Only two other factors correlated significantly with freshman GPA. When entered into a step-wise regression analysis with SAT scores and high school average, the academic achievement factor entered first (zero-order $r = .45$ for males and .41 for females). The zero-order correlation was greater than the multiple correlation of SAT-V, SAT-M, and high school record for males, which was .40. When all variables and factors were combined, the multiple correlations yielded were .56 for males and .53 for females, an incremental increase of .16 for males and .09 for females

beyond that attained using the traditional predictors of test scores and grades.

Richards, Holland, and Lutz (1967)

This study contains analyses of extensive data from 35 colleges in the ACT assessment program on predictors of nonacademic performance in college. Modest correlations between scales on the ACT Student Profile Section and inventory of high school accomplishments and other data) and specific college achievements in leadership, music, and speech and drama, art, writing, and science are reported for 3,267 males and 2,428 females who were college freshmen in 1966. An eloquent argument is made for the use of such scales on the grounds that they are important in their own right, that traditional measures do not predict nonacademic college achievements at all, and that quite low correlations yield impressive success ratios when selection ratios are low. It is also argued that multiple assessments for the prediction of multiple criteria are more appropriate than a single assessment for the prediction of a single criterion.

Schaefer (1969)

A 165-item biographical inventory was administered to 800 high school students from 10 schools in the New York metropolitan area. Students who had demonstrated creative accomplishments in art, science, and writing were identified through teacher nominations and other procedures. Standard personality and artistic measures were also made. Control subjects were also identified by teachers as having shown no evidence of creativity. The biographical inventory was shown to better discriminate between experimental and control groups than the traditional personality and artistic scales. Multiple and partial correlations were also computed for the biographical inventory as a predictor of artistic achievement for boys ($R = .64$) and writing achievement for girls ($R = .55$).

Schultz and Green (1953)

This study was a continuation of the study of Myers and Schultz (1950). The 1953 class of the same women's Eastern liberal arts college was followed after having been administered a 145-item questionnaire covering motivation for attending college, intellectual interests, teacher relations, and study habits. An achievement index, based on the residuals of a regression of college performance on academic aptitude test scores, was used to key the items. Modest correlations were obtained for the prediction of freshman GPA using three different scoring keys. Multiple correlations and beta weights were also computed for combinations of predictors including SAT scores, high school grade averages, and the questionnaire. It was concluded that little contribution was made by the questionnaire and, accordingly, that the results did not warrant administering the questionnaire for selection purposes.

Scott and Castles (1975)

In a study of high school students, an attempt was made to predict nonacademic accomplishments of high school seniors in leadership, art, music, social service, literature, dramatic arts and science. A criterion measure, the Talented Accomplishments Questionnaire (TAQ) was developed to obtain information on accomplishments in these areas. Biographical data, a motivational measure, and an intelligence test were used as predictors. The biographical data were obtained using a 118-item questionnaire developed at the University of Georgia (see Owens and Schoenfeldt, 1979, for details). Of the total of 23 predictor variables available from these three sources, only 6 predictors (all from the biographical inventory) contributed significantly to the prediction of the total score on the accomplishment questionnaire. For males, only two predictors (both from the biographical inventory) contributed to prediction, and a multiple correlation of only .19 was generated. For females, five biographical inventory predictors contributed and yielded a multiple correlation of .64. For both males and females, the best predictor was scientific-artistic interest, a factor indicating the extent to which the student enjoyed science and laboratory courses and found them relatively easy.

Segal et al. (1981)

A biographical questionnaire was mailed to 400 biologists who had received their doctorates during the years 1960–67. A total of 335 responded, and questions

about predoctoral accomplishments and experiences were related to numbers of papers published and numbers of citations in the *Science Citation Index*. The best predictor of numbers of papers published was the number of predoctoral publications in scientific journals ($r = .41$). The magnitudes of correlations predicting numbers of citations were somewhat lower, but many were statistically significant. It was concluded that potentially creative biologists are often identifiable through predoctoral achievements but that identification at an earlier level of education (college or high school) by this means is difficult.

Taylor and Ellison (1967)

This paper presents a summary of research on the use of biographical information to predict performance and accomplishment in science. The initial inventory used in these studies contained 300 multiple-choice items arranged into four sections: developmental listing up to age 21, parents and family life, academic background, and adult life and interests. The criteria of performance were based on official records at research centers of the National Aeronautical and Space Administration, data on numbers of publications and numbers of patents obtained, and special ratings by supervisors in the research settings.

In the first of three studies reported, 354 scientists who had completed the inventory were divided into two subsamples of 178 and 176. Biserial correlations were then computed for each alternative of each item against each criterion in a double cross-validation design, and validity coefficients of .55 (for predicting an overall criterion) and .52 (for predicting creativity) were obtained. In a second study reported, the creativity criterion was predicted, with a correlation of .47 in a sample of 300 scientists. Follow-up studies of true prediction were less successful, but the authors note that the criterion data were collected from diverse research centers, with little control. Predictive correlations in the follow-up studies ranged from .05 to .44. Examples of biographical items given suggest that the inventory would correlate highly with the usually available academic record data. For example, "What percentage of students in your class did you exceed academically?" Moreover, the validity analyses are not reported precisely enough to allow for further analysis.

Walsh (1967)

This was an experimental investigation designed to evaluate the accuracy of information obtained through interviews and questionnaires. A total of 270 male undergraduates at the University of Iowa were randomly selected from residence halls. These students were then randomly assigned to six subgroups, three of which were experimental groups and three of which were control groups. Information obtained was also compared with records available from the university. One experimental group was given an incentive to distort responses. The results indicated that no method elicits more accurate data than another and that incentives to distort information had little effect on accuracy.

Walsh (1968)

Samples of 300 female and 240 male undergraduates were assigned to experimental and control groups to examine the accuracy of self-reported information obtained by different methods (interview and questionnaire) and under different experimental conditions (incentive to distort responses vs. no incentives). No differences were found for six of eight questions asked by both interview and questionnaire. However, one question about the number of semesters a student had been on academic probation was understated on the questionnaire for females and understated in the interview for males. These results indicate that, generally, biographical information can be accurately collected using both interview and questionnaire methods, but both methods may yield distorted information if the questions asked are excessively personal.

Walsh (1969)

A continuation of a line of investigation reported on in two previous papers by the same author, this study introduced an additional social incentive to self-report distorted information in either an interview or through a questionnaire. A total of 192 male undergraduates in a psychology course were randomly assigned to experimental and control groups. The results demonstrated that the method of collecting data was not associated with the accuracy of information in the samples studied. This result was a corroboration of two previous studies by the same author.

Watson (1965)

A biographical inventory was administered to 96 men in an elementary psychology course at the State University of Iowa. A total of 84 of these students had available aptitude test scores, and those 84 cases were retained for analysis. With the exception of father's educational level, none of the biographical items showed a significant relationship with cumulative GPA. Father's education correlated .32 with cumulative GPA, but when academic aptitude was partialed out the correlation was reduced to .24. The Personal Background Inventory described apparently was not a highly sophisticated instrument, since no mention is made of keying techniques. The length of the inventory was not reported.

Webb (1960)

Two biographical inventory keys, one based on an external criterion and the other on a deviate technique, were developed for predicting freshman grade average at Emory University. The inventory contained 200 items of the multiple-choice type. This inventory was administered to men and women entering Emory University in the classes of 1955 and 1956. The inventory was keyed using the 1956 class and cross-validated on the 1955 class. Significant correlations for predicting freshman GPA were obtained using both criterion keying procedures. Increases in validity over that obtainable with the usual test scores and grades were also reported, but it was concluded that the increment obtained was small and of questionable practicality (the multiple R was increased from .65 to .66 for men and from .59 to .60 for women).

Willingham (1965)

Students entering Georgia Tech in 1960 and 1961 completed a 35-item application blank as part of the admissions procedure. The information from this application blank was used to predict both freshman average and voluntary withdrawal. Freshman average was predicted from the application blank information with a correlation of .48. A more important consideration was the incremental gain in prediction added by the application blank over and beyond that attainable using the usual data available. The high school average correlated .49 with freshman average, and the addition of the application blank to the prediction increased the correlation to .55. However, when either aptitude test scores or achievement test scores were also included in the prediction, the incremental value of the application blank was substantially diminished. Voluntary withdrawal from college was predicted with a biserial correlation of .25, but the items that predicted withdrawal tended to be those over which the student had little control.

Recommendations

Browning (1968)

This article reports on an examination of the relationships between pre-employment ratings of 508 teachers and on-the-job performance evaluations during their first full year of employment in the Montgomery County, Maryland, public schools. A total of 2,221 reference ratings (predictors) were made on six scales: teaching ability, relationships with children, professional relationships, personal qualities, and health. A number of different superiors made the teaching performance evaluations (the criterion) including superintendents, principals, supervisors, and heads of education departments. The overall correlation between the pre-employment ratings and first-year teaching performance evaluations was only .13. However, correlations varied somewhat depending on who made the performance evaluations. The highest correlation (.23) was obtained with evaluations made by 96 last supervisors, while the lowest correlation (−.03) resulted from performance evaluations made by 68 professors of practice teaching.

It was concluded that reference ratings had a low level of validity for predicting first-year teaching performance in this school system. It was also observed that years of teaching experience predicted end-of-year evaluations as well as, if not better than, reference ratings (R = .16 vs. R = .13). It was noted that research on the validity of reference ratings is virtually nonexistent but that available evidence provides little support for their validity as predictors.

Calkins et al. (1974)

Reference ratings were related to a Selection Council Index used in admission to the University of Missouri (Kansas City) Medical School's six-year combined B.A./M.D. program. No information was provided concerning the nature of the ratings. Applicants for three different years—1972, 1973, and 1974—were studied. Correlations between the reference ratings and the Selection Council Index were .38, .38, and .28, respectively, for the three years studied. It is of interest to compare these correlations with those between reference ratings and end-of-second-year performance reported in a later study on the same students (see Calkins et al., 1980). The later study showed that the reference ratings correlated much lower, generally, with performance in the college than they did with the Selection Council Index.

Calkins et al. (1980)

Reference ratings were used in an attempt to predict four types of performance through the second year of a six-year combined B.A./M.D. program at the University of Missouri Kansas City School of Medicine. No information was given concerning the nature of the reference ratings. The four outcomes predicted were second-year cumulative GPA, second-year examination, clinical performance evaluation, and status in program. Nonminority and minority groups were studied separately. For 396 nonminority students who entered during the years 1972–77, correlations were .24, .24, .16, and .16, respectively, between the reference ratings and the four performance measures. For 51 minority students who entered during the same years, the correlations were .29, .11, .34, and .07. Thus, the reference ratings would appear to offer some promise for predicting the clinical performance of minority students. The discussion of the results, however, cautioned that the higher minority correlation may have been the result of a greater range in the performance data, since the medical aptitude test score also correlated higher with minority performance (.51, .54, .45, .43). Accordingly, these data suggest that the aptitude test score predicted clinical performance better than did the reference ratings.

Creager (1965)

Reference ratings were used to predict both Ph.D. attainment and B.A./Ph.D. time lapse for 3,491 people who applied for National Science Foundation graduate fellowships in 1955 and 1956. Seven different fields of graduate study were represented, and subjects were classified into these fields by sex for analysis. For nine distinct groups (men and women were combined for some fields), the median correlations for predicting Ph.D. attainment and B.A./Ph.D. time lapse using reference ratings were .17 and .23, respectively. Median multiple correlations were .41 and .40 (GRE scores and grades, without reference ratings) and .41 and .41 (with reference ratings). Thus the median incremental validi-

ties for the reference ratings were .00 (Ph.D. attainment) and .01 (B.A./Ph.D. time lapse).

Creager and Harmon (1966)

This was a study of the effectiveness and validity of the techniques used in the selection of National Science Foundation predoctoral Fellows. In a follow-up of 1955 and 1956 candidates, selection variables were compared with a number of criteria including doctorate attainment, ratings of on-the-job performance several years subsequent to fellowship application, and counts of citations of published work. Predictors consisted of GRE, GPA, and reference ratings. It was concluded that the GRE was most useful for predicting doctorate attainment and citation counts, but that the reference ratings were most useful in predicting later on-the-job ratings.

Cuca et al. (1976)

In their review of medical school admissions procedures, the authors report on studies of the usage of recommendations but no validity evidence. In one survey described, 534 medical school admissions committee members were asked about the value of premedical advisers' reports (recommendations). These recommendations were perceived as most valuable for appraising scholastic achievement and for comparing medical school applicants from the same premedical college. Recommendations were seen as least valuable for assessing motivation for the study of medicine and for assessing suitability as a potential physician.

Farley (1978)

A total of 1,194 letters of recommendation, which were used for faculty appointments in a number of fields at a major university, were examined for content. Sex of applicant and sex of recommender were related to content. There were 250 applicants for positions in 13 different disciplines between 1949 and 1977. For the women applicants, 74 percent of their letters were written by men, and 18 percent by women. The sex of the writer could not be determined for 8 percent of the letters. For the earlier letters, male letter writers for women applicants mentioned marital status and physical appearance of the applicant much more often. However, the study showed that mention of marital status and physical appearance declined markedly over time with only 12 percent of letters received during 1976–77 mentioning marital status and only 1 percent mentioning appearance. It was concluded that there were no differences in letters written in 1977 by applicant sex or writer sex and that recent letters are largely free from mention of criteria unrelated to job criteria.

Guillemin et al. (1979)

A content analysis was conducted of 243 letters of recommendation written for 96 applicants for a position in a sociology department of a Northeastern university in 1975. There were 33 women and 66 men applicants for the junior-level positions. Content analysis categories were organized according to three areas of employment qualifications: scholarly and organizational competence, research abilities, and social compatibility. Of these three areas of competence, the first was found to be correlated with sex. Men were perceived as having superior organizational and scholarly abilities. The smallest differences between men and women were in social compatibility and congeniality. Although the majority of letters was positive in tone, men were much more likely to be described in superlatives. The study suggested that recommendations put women at a severe disadvantage in competition for beginning-level positions at the institutions studied.

Harmon (1966)

This is a summary of 14 years of research on fellowship selection techniques conducted by the National Academy of Sciences, National Research Council. Described are the reliability of instruments and procedures, the prediction of doctorate attainment, the prediction of on-the-job effectiveness, and candidate differences by field, region, and later employer category. Comparisons are given between reference reports and the GRE Advanced Tests as predictors of doctorate attainment, publications counts, citation counts, and on-the-job ratings. It is suggested that Advanced Tests are superior for the prediction of the first three criteria, but that the reference report is best for predicting on-the-job performance. Some attention is also given to the develop-

ment of the reference report used in the NSF Fellowship Program.

Kryger and Shikiar (1978)

In an experimental design, eight hypothetical letters of recommendation (male or female writer, male or female applicant, favorable or unfavorable letter) were written and sent to different personnel managers. A total of 128 personnel managers were randomly selected, and one of the letters was sent. The favorable letters were identical, as were the unfavorable letters, except that the sex of the writer or the applicant being recommended was changed by modifying first names. The personnel managers were also sent a questionnaire requesting evaluations of the applicant. The results showed that women were rated significantly higher on attitude, quality of work, initiative, leadership ability, and general competence. The better evaluation of women applicants was considered as possibly related to affirmative action programs and to the feminist movement generally.

Lewis (1972)

This paper reports on examinations of two types of recommendations. A first study sampled letters of recommendation written by sociologists for students seeking admission to a graduate sociology department in 1967 and 1968. A second study was of a sample of 142 letters written on behalf of chemists seeking faculty appointments. Content analyses of both sets of recommendations demonstrated a pervasive preference for the social ethic (getting along with others) over the puritan ethic (working hard).

Lunneborg and Lillie (1973)

Letters of recommendation written for 123 students admitted to graduate study at the University of Washington during the period 1963–67 were analyzed for sexist comments about the 38 women in the sample. It was judged that sexist comments were made about 11 women. Only one of the 85 men students received a sexist comment. These sexist comments were considered derogatory even when the general tone of the letter was positive. The authors suggest that sexist comments of the type noted contribute to the imbalance in male–female graduation rates, though no data are presented showing how the sexist comments relate to admission decisions.

McCormick and Ilgen (1980)

In reviewing the use of references in personnel selection, the authors state that at least four requirements must be met by those serving as referees: adequate observation of the individual in relevant situations, competence to make the necessary assessments and evaluations requested, willingness to give frank opinions, and capability to express themselves so that the recipient interprets evaluations as intended. It is suggested that there is abundant evidence that these requirements are usually not met by referees. Questionnaires are offered as one way of overcoming the deficiencies of conventional letters of recommendation. But evidence is cited that even questionnaires fail to elicit information that is useful in predicting future performance. The authors conclude that references are probably used most commonly in a negative fashion to reject candidates, but that they can be used positively to make affirmative decisions when some of the referees' requirements are met.

Mehrabian (1969)

A sample of 79 students enrolled in the graduate psychology program at UCLA was studied. A number of measures was available for each subject, including ratings by graduate school faculty of the psychology department that the subject had attended as an undergraduate. The overall promise of a candidate was rated on a 10-point scale. Additionally, each candidate's research orientation was assessed from his biographical statement in combination with his letters of recommendation. Dependent variables consisted of professors' evaluations of graduate achievement, graduate grades in content courses, graduate grades in statistics courses, and an overall graduate school performance index. Zero-order correlations between the overall promise ratings and the first three criteria were .25, .25, and .38, respectively. Zero-order correlations between the research orientation rating and the first three dependent variables were .19, .31, and .06. Thus, for example, the overall promise ratings tended to predict statistics course grades better than did the research

orientation rating. Both zero-order and multiple correlations were made between selected predictors and the overall graduate school performance index. The baseline used for the step-wise analysis was a Graduate Record Examinations–Miller Analogies Test combined predictor, which correlated .43 with the performance index. Adding the overall promise rating yielded a multiple R of .51 (indicating an incremental validity of .08 for the recommendation). However, it is noteworthy that undergraduate GPA was not included as part of the baseline prediction.

Mendel and Tabb (1977)

It is reported that evaluations by advisers of predental students have more value for selection committees if the adviser has had the student in his class or has worked closely with him in other ways. In some colleges, predental advisers act as an unofficial screening committee for a student's entrance into dentistry, and accordingly they are in a position to encourage undergraduates to apply for admission or not. It is noted that these evaluations have not been correlated with success in dental school or after graduation but are very useful in recruiting.

Mosel and Goheen (1958)

The relationship was studied between reference evaluations of job applicants submitted in an employment recommendation questionnaire (ERQ) and supervisors' ratings after employment. Average total scores based on the entire ERQ showed very low or no validity, even though ratings of character and occupational ability had small relationships with supervisors' ratings. Performance of 1,193 civil-service employees in 12 skilled jobs was compared with scores derived from the U.S. Civil Service Employment Recommendation Questionnaire (ERQ). For the 12 skilled jobs, predictive correlations between the ERQ and supervisors' performance ratings ranged from −.10 (for machine operator) to .29 (for radio mechanic). It was concluded that the ERQ had practically no value for the prediction of later supervisory ratings, but it was noted that the findings did not preclude the possibility of developing a more effective ERQ through empirical methods.

Mosel and Goheen (1959)

An analysis is reported of 795 Employment Recommendation Questionnaires (ERQs) on 400 applicants for six professional jobs and one trade in the federal service. The dependent variable was a 10-factor performance ratings made by supervisors after employment. Correlations were computed between ERQ ratings and supervisors' ratings of five different respondent types: personnel officers, previous supervisors, co-workers, acquaintances, and relatives. Only two of the five types of respondents' ratings showed significant correlations with supervisors' ratings. Previous supervisors' ratings (N = 188) correlated .19, and aquaintances' ratings (N = 182) correlated .20 with supervisors' on-the-job ratings. Combinations of previous supervisors' ratings and personnel officers' ratings had significant and positive predictive correlations with some jobs but significant *negative* predictive correlations with other jobs.

Nicholson (1970)

In this long-term study at Brown University, an Admission Index (AI) was developed from a summation of ratings made by secondary school counselors' ratings. The ratings summed were ratings of "motivation" and "promise." A variable similar to AI was constructed for 1950–52 classes. During this early period the practice was to send a rating form to the school principal and to two other raters nominated by the candidate. Raters were asked to judge students as "superior, good, average, below average, or do not know" on nine personality characteristics such as industry and reliability. Criteria predicted were GPA, graduation, advanced degree attainment, and a long-term rating of success made by alumni from the same classes 15–17 years after graduation. Correlations between AI and criteria are reported only for GPA, where predictive relationships between AI and GPA ranging between .22 and .37 were obtained. Generally these relationships were less than those for secondary school rank but greater than those for SAT-V and SAT-M. Tests of mean differences on AI between graduates and nongraduates, advanced degree attainment vs. no advanced degree, and long-term success

ratings were significant for both veterans and non-veterans in the 1950–52 classes.

Quann (1979)

The use and value of written recommendations was significantly affected by the Family Educational Rights and Privacy Act of 1974, as amended. This act provides for students' access to recommendations, although they may waive this right if they choose. These difficulties with recommendations are listed:

- They generally mention only positive attributes.
- They are almost always written by persons selected by the applicant.
- They are often unstructured and difficult to compare.
- Their effectiveness depends on the writing skills of the person who prepared the recommendation.
- It must be assumed that the writer is the person designated on the form.

One positive aspect that was observed is the opportunity the recommendation provides for the emergence of important characteristics not revealed in other procedures.

Reilly (1976)

Analyses were made of the utility of ratings made on the Standard Law School Recommendation Form (SLF) of 119 students admitted to one law school located in the Far West. On the SLF, seven possible ratings (truly exceptional, outstanding, unusual, very good, average, below average, and no basis for judgment) are made of eight different applicant traits (native intelligence, independence of thought, oral communication effectiveness, written communication effectiveness, industry and motivation, judgment and maturity, leadership ability, and personality). Factor analyses of these eight scales suggested the operation of a halo effect, since the eight different judgments all correlated highly with the first factor. Significant zero-order predictive correlation were obtained between law school first-year average (FYA) and several of the ratings. However, when multiple regression analyses were conducted, only oral communication effectiveness added significantly to a multiple correlation based on GPA and LSAT. The incremental validity (increase in R) resulting from the oral communication effectiveness rating was .02.

Richards and Taylor (1961)

One analysis reported in this article was of the predictive utility of ratings made by undergraduate premedical advisers of applicants to the University of Utah College of Medicine. Subjects were restricted to 57 students admitted to the College of Medicine who had taken their undergraduate training at the University of Utah. The Premedical Advisory Committee, meeting as a group, made a single rating of each student on a five-point scale. These ratings correlated .25 with first-year medical school grade averages, .31 with second-year averages, and .03 with third-year averages. In a multiple correlation, the ratings increased prediction of second-year averages from .42 (based on undergraduate grade average only) to .45. No multiple correlations were reported for the prediction of third-year average or for combinations of grades, test scores, and ratings as predictors. In this multiple correlation, the beta weights for undergraduate grades and recommendation ratings, respectively, were .36 and .20, indicating a substantial contribution from the ratings even though the incremental validity of the ratings was only .03. No information is given concerning data available to the Premedical Advisory Committee members when ratings were made.

Shaffer, Mays, and Etheridge (1976)

It was hypothesized that prospective employers would favor job applicants who filed confidential (versus open) letters of recommendation. The question of preference is important because applicants have the right, under the Family Education Rights and Privacy Act of 1974 (Buckley Amendment), to waive their right of access to letters of recommendation at the time the employer requests them. An experiment was designed to test the hypothesis on a sample of 144 men and women applicants, using judges who were cast in the roles of prospective employers. The results suggested that the hypothesis was true (that employers would favor applicants with confidential letters), but the authors cautioned that the experiment was not conducted in an actual employment situation.

Willingham (1974)

In a review of selection procedures for graduate schools, the author observes that recommendations appear to be a poor predictor of whether a student will complete a doctoral program. The evidence on validity, however, does not deter a widespread use of recommendations, despite the time required to prepare them. It is noted that "the Achilles heel of personal references is the unreliability of, or lack of comparability, among judges." Reliabilities of simple references are reported to be in the low .30s, and extensive efforts to improve reliability with multiple ratings have been unsuccessful. Nevertheless, it is concluded that recommendations can be highly relevant for judging a student's suitability for a particular graduate program.

Interviews

Asher (1970a)

Examined a new format for the selection interview called the Q by Q (Question by Question) format, in which the interviewer makes a judgment following each individual interview question. The Q by Q format was developed in an attempt to resolve the problem of unreliability in the typical selection interview. The procedure was to record on television tape a series of 20 interviewers and then to have two separate groups of raters view the tapes. The first group rated using the Q by Q format and the second group made a global end-of-interview rating. The average inter-rater reliability for the Q by Q format was .82 when the task was judgment of probable success in statistics. The same judgment yielded an average inter-rater reliability of only .60 when a global end-of-interview rating was used.

Asher (1970b)

The primary interest in this study was to explore how the physical appearance of applicants influenced the reliability and validity of the selection interview. The subjects were 20 students enrolled in their first statistics course; the selection problem was to predict from an interview success in this course. Two success criteria were used: the final course examination and paired comparison ratings by the instructor. The analyses of inter-rater reliabilities for two groups of raters, those who viewed television tapes of the interviews and those who only listened to the audio part of the interviews, showed little effect of appearance on the inter-rater reliability. A dramatic effect on inter-rater reliability was obtained when a structured interview (the Q by Q format in which a judgment is made after each interview question) was compared to a global end-of-interview rating. Median inter-rater reliability was .82 for the Q by Q structured format and only .58 for the global rating. The validity analyses, however, indicated that viewing the interview resulted in better predictions than did listening only.

Asher (1971)

A structured interview procedure, the Q by Q format, was compared to traditional language aptitude tests as

predictors of language performance after six weeks of language instruction at a Defense Language Institute. Several different languages were involved. The results of a study of 87 students in seven different language groups showed that, when inter-rater reliability was high, so was the predictive validity of the structured interview procedure. However, traditional language aptitude tests yielded higher predictive validity coefficients.

Burgess et al. (1972)

Structured interview ratings were related to admissions decisions at the University of Missouri Kansas City School of Medicine. A total of 73 candidates were interviewed for 40 openings in the first year of a six-year bachelors-medical degree program, and 132 candidates were interviewed for 36 openings in the third year of the program. Both physician and nonphysician interviews were conducted, and each type of interview was structured with specific questions to be asked and specific types of ratings to be made. Most of the ratings made by physicians, however, were different from those made by nonphysicians. These conclusions were suggested from the data.

1. The extent of agreement between physicians and nonphysicians was low, except for the overall decision of whether to accept or reject an applicant.
2. Both types of interviewers tended to form their judgments on the basis of an overall global impression of the applicant.
3. Correlations between physician and nonphysician ratings were low but positive.
4. The influence of both physician and nonphysician recommendations on selection council decisions was low.

Calkins et al. (1974)

Interview ratings were found to correlate highest with a council selection index after admissions policy was changed to a two-stage procedure in which the selection council (the second stage) was unaware of scores and other academic measures used for the first stage. Applicants were those to a six-year baccalaureate-M.D. program at the University of Missouri: Kansas City. Correlations with the selection index were obtained during three years of study—1972, 1973, and 1974—for which the first year represented the old policy and last two years the new policy. In 1972, before the two-stage policy was introduced, race (white = 0, other = 1) correlated −.30 with the council selection index. In 1973 and 1974, after the new policy was introduced, race correlated only .06 and .02 with the council selection index.

Calkins et al. (1980)

Both physician and nonphysician interview ratings were used to predict end-of-second-year performance in a six-year combined B.A./M.D. program at the University of Missouri: Kansas City School of Medicine. Four kinds of performance indicators were used: second-year cumulative GPA, second-year examination, clinical performance evaluation, and status in program. The physician interview ratings correlated .12, −.01, .05, and .09, respectively, with the performance indicators in a sample of 396 nonminority students. For 51 minority students, those correlations were .19, −.03, .25, and .07. Those correlations would suggest that the interview offers promise as a predictor of clinical performance for minorities. However, the discussion of the results cautions that the higher minority correlations may be related to a wider range in the performance evaluations for minorities. Such a relationship is indeed suggested by other results showing that aptitude test score correlated more closely (.45) with clinical performance for minorities than did the interview rating. The aptitude test score correlated only .17 with nonminority clinical performance.

Char et al. (1975)

This study surveyed medical school admissions policies and procedures and noted a universal reliance by medical schools on GPA, MCAT, and the interview. General dissatisfaction with all admissions processes was observed, but divergent views on the utility of the interview were in evidence. The authors explored the team approach to the interview (multiple interviewers) at the University of Hawaii medical school and concluded that the use of a team of interviewers resulted in more interesting and stimulating interviews, helped interviewers to be more objective, helped identify the biases of interviewers, and revealed more of an applicant's personality.

Cuca et al. (1976)

This report contains a very useful review of the interview as used and researched for medical school admissions. The role of the interview in admissions is considered from two perspectives: (1) that of its proponents who argue that it distinguishes better among applicants to very selective institutions than does the MCAT and GPAs, and (2) that of the opponents of the interview who contend that there is no actual proof of its value and who point to the excessive costs in terms of the professional time required for interviewing. The report also includes information on the training of interviewers, especially for minority admissions procedures.

Doering et al. (1979)

This paper notes that the traditional interview used by most professional schools is usually nonproductive in the sense that the observer makes highly subjective judgments on attributes that have no particular relevance to success. The authors believe, however, that a structured personal interview—developed from observations of successful dental practitioners across the country—can provide useful information. A structured interview procedure was validated using the entire senior class (N = 86) at the University of Iowa College of Dentistry. The results showed significant agreement between interview scores and faculty ratings of students in several areas. The structured interview was used to select the 1978 entering class, and a comparison was made of the experimental group and a hypothetical class selected using traditional indicators (GPA, Dental Aptitude Test scores, and number of years of preprofessional education). It was concluded that the two classes were significantly different, since 20 percent of the actual class would not have been admitted without the interview. However, there were no significant differences between the two classes in either the academic or perceptual motor ability mean scores of the Dental Aptitude Test.

Dunnette (1962)

The continued uncritical use of the personal interview offers a clear illustration of excessive delay in undertaking needed research. The author observes that nearly everyone uses this costly, inefficient, and usually nonvalid selection procedure, but that there has been little effective research to support the use of the personal interview as a prediction tool. It is concluded that until more research is done, "its continued use had best be rationalized on the basis of its utility as a public relations device rather than as a personnel procedure" (pp. 291–292).

Dyer (1976)

In this useful monograph, the interview is described as a conversation between two people. Both structured (called standardized) and unstructured (called open) interviews are considered. The author notes the traditional measurement concerns (reliability, validity, etc.) as well as practical matters of importance in interviews. With respect to practicality, it is observed that "assessing the attributes of individuals by means of the standardized diagnostic interview is a lot more expensive on a per-capita basis than assessing the same attributes by means of standardized paper-and-pencil tests administered to individuals in groups of 40 or more persons" (page 14). The author suggests that practicality may be as important as any other attribute of a measure.

England and Paterson (1960)

The authors suggest that there should be a moratorium on books, articles, and other writings about "how to interview" and "do's and don'ts," until there is sufficient research evidence about the reliability and validity of the interview as an assessment device to warrant its use in such work.

Frank and Hackman (1975)

This study suggests that interviewer-interviewee similarity is a real and potentially serious source of bias and validity in selection interviews. Data were collected for three admissions officers at a selective university and for 29 applicants interviewed by each officer. Similarity was judged on the basis of sex, socioeconomic level, urban-suburban, secondary school type, intellectual orientation, orientation toward work, artistic involvement, organizational involvement, and service orientation. Correlations between interviewer similarity to ap-

plicants and favorable bias toward applicants were positive and statistically significant for two out of the three admissions officers. The correlations were strong for one of the admissions officers. Several previous studies of interviewer similarity conducted in employment settings are cited that yielded the same conclusion.

Frederiksen (1954)

Although the interview is the most widely used and is probably the most trusted tool in the kit of the admissions officer, it may be of little value except for the assessment of the most superficial qualities. As evidence, the author cites two studies, one of elect ician's mates during World War II and the other of clinical psychology trainees, in which a single test score was a better predictor of performance than elaborate interview procedures. This outcome was especially interesting in the second study, as the performance used as the criterion included more than grades in the training program.

Gough (1967)

There is little evidence to support the use of an interview in medical school selection. "The admissions interview, although addressed to variables of great concern, seems to possess little validity as a forecaster of performance in training or practice. . . ." However, substantial correlations are reported between personality measures and cumulative four-year GPA in medical school. The study also reports validity evidence for the Adjective Check List as a predictor of good physician characteristics.

Gough and Hall (1975)

A number of procedures, including interviews, were used to predict long-term clinical and academic performance for 661 medical students at the University of California School of Medicine, San Francisco, who had entered in the years 1960–63 and 1966–67. Two criterion factors were developed based on all four years of medical school. The first factor, labeled clinical competence, loaded highest on faculty ratings of clinical competence (.92) and faculty ratings of general medical competence (.89). The second factor loaded primarily on first-year GPA (.90) and second-year GPA (.84). Mean ratings by admissions interviewers correlated modestly (.13) with clinical competence but not at all with the academic performance factor (−.01). Correlations slightly higher were obtained between clinical competence and some Adjective Check List scales. The study suggests that clinical competence, as measured, is not very predictable, but that academic performance is predictable. Premedical GPA and MCAT science scores correlated in excess of .30 with the academic performance factor.

Gough et al. (1963)

In an extensive study of 14 medical school classes at the University of California School of Medicine, San Francisco, a number of predictors and criteria were investigated. Admissions committee interview ratings were correlated with medical school GPA for each of the four years of medical school for classes graduating in 1951 through 1964. For the 1959 class, admissions committee interview ratings were correlated with faculty rating factors in 1958 and 1959. The faculty rating factors were empirically derived factors (through factor analysis) based on student qualities such as clinical judgment, diagnostic skill, and identification with medicine. While the interview as a predictor of medical school grades demonstrated little promise (median predictive correlations of .01, .04, .05, and .05 were obtained with GPA for the four years of medical school), the prediction of the faculty rating factors was more encouraging. The interview rating correlated .07, .10, and .24 with rating factors I, II, and III, respectively. For 1959, only two rating factors were reported. The predictive correlation of the interview rating with the first faculty rating factor was .14 and with the record was .19.

Gough et al. (1964)

Interviewers' ratings and other preadmissions measures were used to predict criterion factors developed from faculty and peer ratings of medical school performance for 13 classes (1951–1963) over four years of study at the University of California School of Medicine, San Francisco. Correlations between the interview ratings and the later faculty rating factors ranged from .05 to .24. The peer ratings correlated less well with the interview ratings—in fact, all relationships were negative. Interestingly, none of the predictors, including MCAT scores and premedical GPAs, correlated

consistently with either the faculty rating factors or the peer rating factors.

Keck et al. (1979)

This was an investigation of clinical performance of resident physicians as judged by a 33-scale Resident Evaluation Form (REF) completed by supervising physicians. The 33 scales were used to assess the performance of residents in areas of professional responsibility, relations with colleagues, self-appraisal ability, clinical skills, relations with patients, work in groups, critical thinking ability, independent learning ability, and problem solving. Each scale measured a specific attribute of clinical performance and was behaviorally anchored at both positive and negative ends. This residence performance criterion was predicted by an interview rating given by a physician interviewer at the time of application to medical school. The interview rating was a composite of the interviewer's ratings of the applicant's perception of a motivation for medicine, interpersonal skill, personal soundness and maturity, decision-making ability, independence, personal ethics, and social awareness. For the 55 residents for whom both the criterion rating (the REF) and the interview rating were available, a predictive correlation of .19 was obtained.

Kegel-Flom (1975)

Intern performance as measured by supervisor, self, and peer ratings was compared with medical school admissions information collected several years earlier for 110 interns in residence at 54 different medical facilities. Interview ratings and several other admissions variables were related to the intern performance ratings. Correlations with interview ratings were low for supervisor ratings (.04), self-ratings (−.04), and peer ratings (.12). Better predictions were made with the SVIB Physician scale (r = .25 with supervisor ratings) and with the CPI Tolerance scale (r = .35 with peer ratings).

Korman et al. (1968)

Sixty-eight applicants to the class of 1962 at the University of Texas Southwestern Medical School (Dallas) were interviewed by 3 faculty members prior to admission. Each faculty member rated each applicant on a total of 23 7-point scales. These ratings were made without reference to MCAT scores and premedical GPAs. The ratings were related to medical school performance information collected during the student's senior year in medical school and to internship ratings completed by supervisors at the end of the internship year. General achievement (including clinical performance during the first four years of medical school was significantly predicted by ratings of "Poise" (r = .30) and "Acceptability as a student" (.22). Peer esteem, based on ratings made by all other senior medical school students of each other, also was predicted by "Poise" (r = .25). Scientist potential was predicted by a number of ratings: "theoretically oriented" (r = .39), "intellectual curiosity" (r = .37), and "intelligence" (r = .33). Internship success and humanism, however, were predicted significantly by only one interview rating each, and in general the magnitude of the predictive correlations was low.

Linn and Winograd (1969)

All applicants to the New York University Law School for the years 1965, 1966, and 1967 who had been placed in a "questionable" category (as opposed to "acceptable" or "unacceptable") were interviewed, and ratings were made of their probable performance in law school. Interviewer ratings were made on a three-point scale: 3 (accept), 2 (waiting list), or 1 (reject). Correlations between these ratings and law school GPA were: .00 with first-year GPA, −.05 with second-year GPA, −.02 with third-year GPA, and, −.03 with law school cumulative GPA. Sample sizes were 332 for first-year GPA, 177 for second-year GPA, 83 for third-year GPA, and 81 for cumulative GPA. It was concluded that "if the sole purpose of interviews is to improve the predictive accuracy for applicants in the questionable category then the results of this study provide no justification for the interview procedure."

Litton-Hawes et al. (1976)

This was a study of the training of admissions interviewers at the Ohio State University College of Medicine. Five admissions committee members were videotaped during three admissions interviews, and their verbal and nonverbal behaviors were classified. Com-

mon problems identified were reliance on the applicant's written file and inefficient use of time. An instructional videotape (for training admissions interviewers) was developed and is reported to be in widespread use both in the U.S. and abroad.

Mayfield (1964)

A review of research on the selection interview as used in employment settings (primarily) between 1915 and 1963, it concludes with 15 generalizations about what is and is not known about the selection interview. Some of those generalizations are:

1. General ratings based on unstructured interviews with no prior information have extremely low inter-rater reliability. That is, the interview as normally conducted in a selection situation is of little value.

2. In an unstructured interview, material is not consistently covered by all interviewers.

3. When interviewers do obtain the same information, they are likely to interpret or weight it differently.

4. Structured interviews, in general, provide a higher inter-rater reliability than do unstructured interviews.

5. Although the reliabilities of interviews may be high in given situations, the validities obtained are usually of a low magnitude.

6. The only trait or characteristic estimated reliably and validly from interviews is intelligence or general ability.

It is concluded that "the interview as normally conducted in a selection situation is of little value." (p. 249)

Mendel and Tabb (1977)

This paper covers a broad range of problems in admissions to dental schools. The factors used to identify the most qualified applicants are described. Interviews are seen as very useful for recruiting but not for selection.

The authors observe that whether to interview candidates for admission to dental school is a controversial subject and that 40 percent of all schools interview over half of their accepted applicants. They conclude that the interview has little effect on the admissions decision: "It expends time and money for the applicant and for the faculty of the dental school. The interview is, however, a valuable recruiting instrument."

Milstein et al. (1981)

This study compared two groups of students who applied to the Yale University School of Medicine for the class of 1976 but attended other medical schools. The first group consisted of 24 applicants who were interviewed and accepted. The second group was made up of 27 applicants who were interviewed and rejected. Both groups attended the same 10 medical schools. Performance on National Board of Medical Examiners (NBME) examinations and dean's office evaluations were used to compare the two groups. No significant differences were observed in 13 different NBME tests and 6 different dean's office evaluations. During the admissions year for the class of 1976, members of the admissions committee devoted more than 1,600 man-hours to interviewing and deliberating over some 700 applicants during the second stage of the admissions process.

It is noted that the cost, generally, for the admissions process to medical schools nationally is about $110 per applicant. Given these high costs and the apparent ineffectiveness of the interview, the authors conclude that perhaps the fairest and most effective admissions procedure would be an initial screening followed by a lottery selection.

Palacios et al. (1966)

This paper describes a study of the interview as a predictor of occupational achievement for blind job applicants. Interviews with 144 blind adults were objectively and quantitatively scored. The highly structured interviews consisted of 43 open-ended questions designed to assess adjustment variables in a blind adult population. Interviews were tape recorded and later transcribed to typescript. Each interviewer response was carefully scored and weighted for the degree of maladjustment indicated. Seven different predictive scales were developed and were then correlated with job success as interpreted by a job hierarchy scale. The multiple correlation of all seven scales with the job hierarchy scale was .48, even though one of the seven scales did not correlate at all with the job hierarchy scale.

Richards and Taylor (1961)

As part of a larger investigation, a study was conducted to determine whether interviews increased accuracy of

prediction of medical school performance over what could be achieved with undergraduate achievement and test scores. Subjects were 65 students admitted to the University of Utah College of Medicine in 1955, 1956, and 1957. These students were applicants who had been classified in a "middle group" for whom more information was evaluated prior to acting on their applications. Applicants in this middle group were interviewed by groups consisting of two to four members of the admissions committee. Each of the interviewers independently rated the applicant on three scales: (1) personal characteristics, a four-category scale; (2) chances for success in medical school, a four-category scale; (3) recommended decision, a six-category scale. Correlations of these scales with first-, second-, and third-year medical school performance ranged from .15 to .32. Multiple correlations and beta weights were also developed, and these indicated that the interview added to the prediction of first-year average over what was obtained with only grades and test scores (R increased from .37 to .40).

Smith (1965)

A lengthy (67 items) structured interview was used to explore differences between 31 high-aptitude nonachievers and 32 high-aptitude achievers among 154 freshmen at the University of Kentucky. The interview covered socioeconomic background, high school background, attitude toward authority, personal needs and aspirations, academic adjustment, peer culture, and satisfaction with and perception of the university experience. A number of differences were found between the achievers and nonachievers, including these: more nonachievers came from metropolitan areas, more achievers were affiliated with Protestant churches, achievers had performed better in high school, achievers had better study habits, nonachievers were more concerned with status, achievers participated in a greater variety of activities, and achievers were satisfied with their choice of college major. This interview study illustrates the observation that highly structured interviews tend to take the form of biographical inventories.

Splitstone (1974)

The author observes that the interview may be for information giving or for evaluation, and that colleges use the interview for either one or the other and frequently for both. One New England college requires an all-day visit/interview on campus, which serves as the application. Problems and utilities of alumni interviewing are described—especially their value in creating good relationships with prospective students in local communities.

Ward (1955)

The article describes experiences with interviews in the admissions office at the Harvard Business School, where each applicant (at the time) received two or three relatively unstructured interviews. Ward reports that correlations between ratings of the same applicants yielded reliability coefficients of .6, despite many variations in technique and content among interviewers. An experimental investigation is related in which one group of applicants in one year was admitted on the basis of other information available. A comparison of the performance of the two groups in the first year of the MBA program revealed no significant differences in performance for the two groups. Ward states that, as a result of this experimental investigation, the school discontinued use of the personal interview in admissions for a trial period.

Weinstein et al. (1979)

This study reports that despite the almost universal use of the interview as a selection technique, the "one-to-one," "face-to-face" interview has seldom proved to be effective. Interviews tend to confirm the prior impressions of the interviewer, or, if changes do occur, they tend to be disproportionately affected by negative information. The authors suggest that where no prior information is available to the interviewer, decisions tend to be made within the first 30 seconds of the interview, and often these decisions depend heavily on the quality of the previous interview.

Wright (1969)

This is a review of studies of the interview classified as microanalytic studies, macroanalytic studies, and re-

lated studies. It is reported that the structured interview continues to show strong inter-rater reliabilities and that there is some evidence that it is being increasingly used as a selection and promotional tool. But it is emphasized that "there is still little evidence as to the efficacy of the interview, either in terms of cost or of optimum length." The author observes that there has been a disturbing lack of research on the selection interview, even though one finds numerous articles about it. One survey is reported in which 93 percent of the companies surveyed had made no systematized attempt to validate their interviewing programs despite their overwhelming preference for the interview as a selection tool. Recommended for future research are more studies investigating the validity of the structured interview.

Interests

Ayers and Rohr (1974)

Overall quality point average (QPA) and total scores on the National Teacher Examination were the criterion measures for various predictor variables for 603 graduates of a teacher preparation program. Three scales of the Kuder Preference Record (Vocational and Personal) were used as predictors. These scales were only slightly related to successful college work, and five out of six correlations were negative.

Baird (1969b)

The Vocational Preference Inventory (VPI) was administered to approximately 12,000 college freshmen in 31 institutions during the spring of 1964. College accomplishments for these students were obtained in the spring of 1965 toward the end of their sophomore year. Complete follow-up data were collected for about 5,000 of the students. The VPI was scored for 11 scales: realistic, intellectual, social, conventional, enterprising, artistic, status, aggressive, masculinity-femininity, infrequency, and acquiescence. None of these scales contributed importantly to multiple predictions of either academic or nonacademic outcomes. The results suggested that interest inventories are less effective as predictors of college accomplishments than are past activities, goals, and self-descriptions.

Banreti-Fuchs and Meadows (1976)

In this study of 114 males and 125 females enrolled in introductory psychology, the students were selected as either high, average, or low achievers. Items on various aptitude, interests, and personality measures that discriminated among the three achievement levels were subjected to factor analysis. All subjects completed the SVIB for men or the SVIB for women. Seven factors for males were indicated and two for women, but none loaded heavily on SVIB variables. The authors concluded that results "underline the modest role of interest variables as a factor in discriminating between achievement levels of males and the nearly irrelevant factor for women" (p. 217).

Belcastro (1975)

Scores on the SVIB and other measures were used to discriminate between graduating seniors who did or

did not complete their teacher training program. For males (N = 211) and females (N = 84) interest predictors contributed less than personality predictors in the discrimination. It was noted, however, that a cross-validation with larger samples would be needed to confirm these findings.

Chen et al. (1967)

The Kuder Preference Record was one of several measures studied, through factor analysis, as predictors of GPA at the University of California School of Dentistry during the 1964–65 academic year. Two factors were derived for the several admissions variables. The Kuder scores did not load significantly on either factor, suggesting that interests, as measured by Kuder scores, are not important influences on dental school performance.

Harrington (1969)

Freshman GPA for 746 men who entered Ohio University in the fall of 1966 was correlated with several predictors, including the SVIB achievement scale. Of all predictors, the next to smallest correlation (.29) with GPA was using the SVIB achievement scale. Other predictors consisted of ACT test scores, high school rank, and two scales of a biographical inventory. In a stepwise multiple correlation, the SVIB was the last significant variable to enter, and the multiple R was increased only from .528 to .532.

Keck et al. (1979)

Long-term performance in medical training was related to several medical school admissions variables. For 56 interns who were in various stages of residency, a Resident Evaluation Form (REF) was sent to their supervisors by the research staff of the medical school from which they had graduated (University of Missouri: Kansas City). Several measures obtained at the time these residents applied to medical school were then related to the REF—a comprehensive set of 33 seven-point scales. One of the admissions measures was Holland's Self-Directed Search (SDS). Correlations between the REF performance the six scales of the SDS ranged from −.23 (for conventional) to .13 (for realistic). When these scales were combined with other predictors in a step-wise multiple correlation, the realistic, artistic, and social scales added significantly to predicting the REF evaluation.

Kegel-Flom (1975)

In an interesting long-range study of medical intern performance as judged by supervisors and peers, a number of medical school admissions measures (including the SVIB) were examined. The physician scale of the SVIB correlated best of all admissions variables with supervisor ratings of intern performance (r = .25). However, the physician scale did not relate well to peer ratings of intern performance (r = −.07). No multiple correlations of the combined scales of the SVIB with intern performance were reported, but patterns of correlations would suggest reasonably good prediction using multiple scales of the SVIB.

Lavin (1965)

In this comprehensive review it is observed that interest measures have generally been put to three uses in the prediction of academic performance. The first use is to attempt to improve the predictions of grades under the assumption that students interested in particular courses will get better grades in them. A second use is in the prediction of academic performance from interests alone. A third use of interest measures is to gauge clarity, intensity, and other levels of interest. Several studies are reported that suggest that interest measures do improve predictions made from academic measures and that intensity and clarity of interest is related to academic performance.

Lindsay and Althouse (1969)

The academic achievement scale (AACH) of the SVIB and the motivation for grades (MG) scale of the College Student Questionnaire (CSQ) were related to grade point average of college freshmen (299 men, 89 women). The predictive correlations for the AACH were .27 for men and .29 for women. Predictive correlations for the CSQ were .10 for men and .25 for women. When SAT scores and high school grades were included in a multiple regression, MG added .04 to the multiple R for men and .03 for women. The increment added by the AACH was much less (.004 for men and .011 for women). The authors concluded that "whatever variance in achieve-

ment is related to AACH scores, it is almost equally measured by aptitude and past performance" (p. 491). It was reported that this lack of incremental validity for the AACH agrees with Campbell's (1966) study.

Nichols (1966)

A portion of a sample of National Merit Scholarship participants was used to empirically derive scales for predicting both college grades and extracurricular activities. The Vocational Preference Inventory (VPI), among other measures, was used to develop scales. These scales were then cross-validated on the remainder of the National Merit sample and on a normative sample representing high school students nationally. In the National Merit sample, VPI grade scale correlated positively (.11) for men but negatively (−.12) for women. Similarly, the VPI extracurricular scale correlated positively (.18) for men but negatively for women (−.06). In the normative sample, the VPI grade scale correlated positively for both sexes (.10 and .13), as did the extracurricular scale (.22 for men and .07 for women). It was concluded that high school grades were the best predictors of college grades but that nonintellective predictors such as interests and personality were better predictors of extracurricular activities.

Nichols and Holland (1963)

In a comprehensive survey of National Merit finalists, over 1,000 students were asked to rank their preference for six groups of occupations. Each group consisted of six occupational titles selected to represent the six interest keys of Holland's Vocational Preference Inventory (VPI): realistic occupations, intellectual occupations, social occupations, conventional occupations, enterprising occupations, and artistic occupations. These ratings were then correlated with several nonacademic college achievements. Substantial correlations were obtained for the prediction of freshman achievement in literature (.40), graphic arts (.40), and Ph.D. aspiration (.38). Those relationships suggest that a simple rating may be about as good for prediction as a complex and lengthy interest inventory.

Nicholson (1970)

In this long-range study of performance of Brown University graduates, occupational performance for a group of risk students (i.e., those with relatively low aptitude test scores) was related to scores from the SVIB and the Occupational Interest Inventory (OII). Data were obtained 15 to 17 years after graduation for Brown classes of 1950 to 1952. Job classifications from alumni records were coded according to the *Dictionary of Occupational Titles* and then assigned levels of responsibility and skill. The sample was then classified into three occupational categories: science/technology, general culture, and business. The validity of the SVIB and OII for predicting long-term occupational choice was supported. Although the interest inventories did generally discriminate between most- and least-successful groups, they were not recommended for use in admissions because of the effort required to administer and interpret them for large numbers of students and because they were less effective than other predictors.

Norris and Katz (1970)

In this two-part study Part I summarized the procedures used in developing an interest inventory, Academic Interest Measures (AIM). Part II presented the results of analyses of follow-up data on the prediction of senior-year and freshman college grades; the prediction of interests in senior year and in college; the structure of abilities, interests, and values; and the relationship between the predictors and occupational criteria for a sample that did not continue in an educational institution. One of the major conclusions of this study was that a simple rating of interest did as well as a predictor of performance and interests in college as did AIM.

Ayers and Rohr (1974)

The Survey of Study Habits and Attitudes (SSHA) was used to predict performance in teacher-training programs for elementary and secondary school teachers at Tennessee Technological University. A total of 203 women in elementary education and 400 men and women in secondary education was studied. All had completed the SSHA as freshmen, and overall GPA for four years of college was related to the SSHA and other variables. Although the SSHA had a positive relationship with GPA for the elementary program ($r = .12$) the relationship was negative for the secondary program ($r = .55$).

Banreti-Fuchs (1975)

A total of 222 men and women introductory psychology students at Laurentian University (Canada) took the MMPI and the Student Interest and Attitude Study (SIAS). Students were then assigned to high, average, and low achievement groups on the basis of introductory psychology course grades. Seventeen items of the SIAS and seven scales of the MMPI were reported to have discriminated significantly among these achievement groups. Eight factors were developed that discriminated among the groups. The first, accounting for 27.5 percent of the variance, was based on seven MMPI scales. The second factor, accounting for 11 percent of the variance, reflected attitudes toward duty and responsibility. High achievers were more likely to attend lectures regularly, keep up with reading material, and finish assignments properly.

Banreti-Fuchs and Meadows (1976)

Subjects in this study were 114 men and 125 women students enrolled in an introductory psychology course. They were administered the MMPI and the Student Interest and Attitude Study (SIAS), among other measures. Subsequently the students were classified into high, average, and low achievement groups on the basis of GPA. Item analyses were then conducted for all instruments to determine which items discriminated among the groups. For men, twelve items from the MMPI and six from the SIAS were found to discriminate significantly among the groups. For women, four MMPI items and six SIAS items discriminated significantly. Factor analysis revealed a first factor for men (accounting for 21 percent of the variance) was related to attitudes toward academic achievement and early mother–child relationship. The first factor for women (accounting for 31 percent of the variances) was related to expectations for academic success and self-perception of intelligence.

Brown and Dubois (1964)

The MMPI and the Survey of Study Habits and Attitudes (SSHA) was related to freshman GPA for 105 students enrolled in the College of Sciences and Humanities and 190 students enrolled in the College of Engineering at Iowa State University during the 1961–62 academic year. For the science and humanities students, three scales of the MMPI and one scale of the SSHA correlated significantly with GPA. For the engineering students, the same three scales of the MMPI and all scales of the SSHA correlated significantly with GPA. Multiple correlations suggested that both the MMPI and the SSHA added small amounts of unique variance to the traditional predictors (test scores and high school grades).

Chen et al. (1967)

Seventy-two second-year dental students participated in this study to predict fifth-semester cumulative grade-point average (GPA). Thirty-three variables were collected in all and included: the 18 subscales of the California Psychological Inventory (CPI), predental GPA, Dental Aptitude Test (DAT), dentist scale of the Kuder Preference Record, manual skills test of the California Performance Test (CPT), and dental GPA. Two of nine factors extracted had significant loadings on dental GPA. The first factor reflected primarily manual skills, and the second factor reasoning skills as measured by verbal and quantitative tests. It was concluded that the CPI was practically useless as a predictor of student performance in dental school.

Donnan (1968)

Two academic variables (SAT and high school rank) and one noncognitive variable (Opinion, Attitude, and Interest Survey, OAIS) were obtained from two groups of freshmen (1962 and 1965) at the University of North Carolina at Chapel Hill to predict GPA. Of the six OAIS

scales (achiever personality, intellectual quality, creative personality, social adjustment, emotional adjustment, masculine orientation) only the first was significantly related to GPA (.41 and .36 for both groups). Intellectual quality predicted GPA for only one group (.23). When GPA was regressed on the academic variables and the six scales of OAIS (R = .63 and .59), only achievers' personality contributed significantly to predictive accuracy over and above SAT-V, SAT-M, and high school rank (8 percent additional variance).

Dunham (1973)

Various measures of achievement motivation were collected from 161 male and 142 female college freshmen and used to predict grade-point average (GPA). Ego achievement score from the California Test of Personality, need achievement scores from the Edwards Personal Preference Schedule, and the Thematic Apperception Test (TAT) were significantly related to GPA. However, extensive missing data in the TAT analysis makes its prediction of GPA suspect.

Edwards (1977)

Adult workers (231 females, 224 males) and high school seniors (N = 237) rated their peers on 16 personality traits from Cattell's list. Significant nonzero correlations with grade-point average of the high school student were found for all traits but two: solitary–sociable and not emotional–temperamental. For the workers the criteria were supervisor rating and mean pay differential. These correlations were reduced by multivariate analyses to yield composite variables of rules orientation, dependability, and loyalty–committed. These variables, along with sex, age, I.Q., educational level, socioeconomic status, and job status were regressed on supervisor ratings yielding a multiple R of .89. The three personality variables made the largest and most significant contribution to the regression equation. It was concluded that the three dimensions of behavior demonstrated high test-retest and split-half reliability and that they postdicted the criterion variables as well.

Entwistle (1972)

This is a review of research on academic attainment using personality measures developed by Eysenck and Cattell. The research is classified into two main categories: school-level attainment and college-and-university-level attainment. It is reported that convincing evidence exists of the superiority of the introvert in higher education and that empirical results indicate that good study habits are related to stable introversion. However, it is emphasized that much of the research on personality and academic attainment is contradictory.

Gough (1967)

This paper reviews procedures used in medical school admissions and makes a case for using personality measures in addition to traditional procedures. Studies of physicians and medical school students are described that indicate that the MCAT and previous grades, when used as the primary selection devices, produce physicians and students who are not widely respected by their peers. It is suggested, therefore, that the usual methods for assessing personal qualities are inadequate.

Gough and Hall (1975)

The Adjective Check List was used to predict long-term medical school clinical and academic performance criteria at the University of California School of Medicine, San Francisco. Correlations in odd and even samples of 331 and 330 students ranged from −.18 to +.21 with a clinical performance factor and between −.11 and +.12 with an academic performance factor. There was little consistency across samples, except for the self-confidence scale of the ACL, which correlated .11 in the odd sample and .12 in the even sample with the clinical performance factor. Multiple correlations showed that addition of the ACL tended to suppress rather than increase predictions. It was concluded that measures of personality offer promise as forecasters of clinical competence, although results to date fall short of a satisfactory level of predictive accuracy.

Heath (1977)

This was a longitudinal study of 68 randomly selected graduates of Haverford College. These graduates were first studied as freshmen in college, then as upperclassmen, and finally in their early thirties. It was reported that traditional predictors and college performance itself were not found to be related to measures of adult maturity and competence. College measures of matu-

rity included the MMPI, the Rorschach, faculty and student judgments, self-ratings of maturity, and ratings of maturity by friends. In their early thirties the subjects completed a questionnaire measuring their maturity and assessing all previous measures of maturity and adjustment, and men were rated by wives, friends, and occupational colleagues on their psychological health. Measures of adult competence—including salary levels, marital competence, parental competence, and interpersonal competence—were also obtained. It was concluded that one of the best predictors of adult competence is the maturity of students when in college as measured by the diverse tests and judgments used in the study. No precise analyses or data are given, however, in support of those long-range predictions.

Himelstein (1965)

Freshmen (N = 281) scores on items of the Minnesota Multiphasic Inventory (MMPI) were combined and used to predict first-semester grade point average (GPA) and ACT cognitive score. Five of the seven subscales were significantly correlated with ACT. As a result of these relationships, the author suggests that the noncognitive predictors may not be independent of intellectual factors.

Keck et al. (1979)

A long-range study of 56 resident medical interns is reported in which admissions variables, including the Omnibus Personality Inventory (OPI), obtained at application to a six-year B.A.-M.D. program at the University of Missouri (Kansas City) School of Medicine were related to a comprehensive Resident Evaluation Form (REF). The REF contained 33 seven-point scales for rating resident performance. Each scale measured a specific attribute of physician performance. Residents varied in their year of residency training: 6 percent were in the fourth year and 45 percent had completed at least six months of the first year of residency. Thus, the time between the collection of predictor and criterion data ranged from about 7 to about 10 years. Three scales of the OPI related significantly with the REF evaluations. (Correlations ranged from .13 to .24.) It was reported also that the OPI scales enhanced the prediction of clinical performance over and above that possible even with six-year medical school performance data and National Board of Medical Examiners test scores.

Kegel-Flom (1975)

Supervisors and peers rated 110 resident interns located at 54 different medical facilities in the United States. These ratings were correlated with a number of preadmission measures obtained as part of the application process at the University of California School of Medicine in San Francisco. One of the preadmission measures was the California Psychological Inventory (CPI). Significant correlations were obtained between the tolerance scale and peer ratings of intern performance ($r = .35$) and between self-ratings and dominance ($r = .30$) and flexibility ($r = -.23$). However, none of the CPI scales correlated significantly with supervisors ratings of intern performance.

Korman et al. (1968)

This was a longitudinal study of 68 medical students followed from their first year to the end of their internship. During the first year the California Personality Inventory (CPI), the Edwards Personal Preference Schedule (EPPS), and a self-rating personality inventory were administered. The scores on these scales served as predictors (among others) of performance. Information on the students' performance was collected during the senior year (e.g., grades, faculty ratings, and peer ratings) and at the end of their internship (e.g., supervisor ratings). Interpretation of the factor structure of these ratings resulted in five factors: achievement in clinical medicine, internship success, peer esteem, scientist potential, and humanism. Predictive validities of scores assigned to these factors were reported. Generally, peer esteem and scientist potential had the largest correlation coefficients with the personality measures. No significant coefficients were found when humanism was predicted. Only self-ratings of personality predicted internship success to any extent. Interestingly, eight significant negative correlations between peer esteem and CPI scales were found.

Lavin (1965)

After a comprehensive and lengthy review of the literature, the author concluded that the relationships between personality measures and academic success were tenuous at best, and that the state of knowledge regarding the relation between personality variables and achievement was still so tentative that it could not be used confidently for practical purposes such as college admissions.

McCausland and Stewart (1974)

The Survey of Study Habits and Attitudes (SSHA) was administered to 154 college freshmen enrolled in an introductory psychology course. Scales of the SSHA were then related to first-semester GPA. For seven scales of the SSHA, predictive correlations ranged from .19 to .40. The highest correlation was for a combined scale, study orientation, which is a combination of the study habits scale and the study attitudes scale. However, the addition of the SSHA scales to high school average and ACT composite score had no appreciable effect on improving the prediction of GPA.

Messick (1967)

The author argues that predictive validity alone is not sufficient in making practical decisions involving personality measures and college performance and notes that personality tests and even personality testers may operate differently under different circumstances. The difference between empirical information about a personality test gathered in a research context versus a presumed college selection setting makes knowledge of response style an important factor in evaluating personality measures. Also discussed were the ethics of using personality measures—e.g., the value judgment involved in selecting a certain test and the determination of an appropriate criterion measure, since different personality and criterion domains are more relevant to some aspects of selection than to others.

Nichols (1966)

Empirical scales for the prediction of college grades and extracurricular activities were derived from a portion of a sample of National Merit Scholarship participants, and these scales were then cross-validated on the remainder of the National Merit sample and on a normative sample of high school students. Across the two samples and two sexes, the average correlation for predicting freshman grades with CPI was .30. For predicting extracurricular activities the average correlation was .17. An Objective Behavior Inventory (OBI) was also used, and empirical scales were developed as for the CPI. Average predictive correlations for the OBI were .23 for predicting grades and .22 for predicting extracurricular activities. It was also reported that these nonintellective scales added to the prediction of grades in a regression equation including high school rank and test scores.

Nichols and Holland (1963)

Several personality scales were related to first-year academic and nonacademic performance of a sample of more than 1,000 National Merit finalists who graduated from high school in 1960. Included were the CPI, Myers-Briggs Type Indicator, a superego scale, a dogmatism scale, an acquiescence scale, a sense of testing scale, a tolerance of ambiguity scale, scales from Cattell's 16 personality factors, and other scales. Zero-order and multiple correlations were conducted for the prediction of GPA, leadership, artistic, dramatic, scientific, and other achievements in the first year of college. It was concluded that the personality measures were second only to actual achievement in high school as predictors of college performance.

Petry and Craft (1976)

A total of 52 students characterized as "high-risk, disadvantaged, and low-income" and enrolled in a teacher education program at Memphis State University was administered the Thorndike Dimensions of Temperament (TDOT) scales and the Study of Values (SOV) scales. These scales were then related to cumulative GPA in the teacher education program. Although no significance tests are reported, correlations in excess of .30 were obtained for two of the personality scales. Multiple correlations combining scales within TDOT and SOV ranged from .28 to .54. When added to the aptitude test score in multiple correlations both TDOT and SOV appeared to add substantially to the prediction, but no significant tests and no cross-validations were reported.

Richards et al. (1974)

This study investigated the prediction of success in a six-year combined undergraduate and medical education program. Predictors included previous academic performance; tests of ability, vocational interests, and personality; evaluation by two interviewers; measures of personal development and participation in cultural and community activities; and attitudes and plans with respect to various kinds of medical careers. The results from three-predictor step-wise multiple correlations indicated the best predictor for each variable tended to be previous standing on that variable. Measures of personality and vocational interest made little contribution to the prediction of most criteria.

Rowell and Renner (1975)

The Eysenck Personality Inventory (EPI) was administered to 136 postgraduate education students in two different programs at the University of Adelaide (Australia), and neuroticism and extraversion scales were compared with academic achievement. Two hypotheses were supported by the results of the study: (1) introverts achieve more highly in structured programs, and (2) extroverts achieve more highly in unstructured programs.

Stein (1963)

The author reviewed the literature reporting relationships between various measures of personality and academic success. Three types of personality characteristics were covered: global, single, and typological. It was reported that global characteristics (e.g., maladjustment, anxiety) distinguish between groups of students of different academic success best when objective rather than projective measures are used. Single personality characteristics (e.g., need achievement, femininity, masculinity) appeared to be best at predicting college success when done in conjunction with other personality variables. Typologies, or clusters of personality characteristics may be used as a single variable for prediction. Another approach discussed is the transactional method, which assumes that college success is a function of the transactional relationship between the individual and his environment. Rather than academic achievement, a standard of performance accepted by others is used, which leads to a hypothetical model of a successful student. The student is then judged to be or not to be congruent with the model. Although it possesses the greatest potential for prediction, the difficulties of gathering data and developing the model currently make this approach the least practical.

Stroup (1970)

Grade point average (GPA), SAT scores, and scores on the California Psychological Inventory (CPI) were collected for college freshman males (N = 968) and females (N = 970). Significant zero-order correlations of GPA with 11 of 18 scales of the CPI were found for both sexes. Stepwise regression, after entry of SAT, indicated that the CPI increased the multiple R for males from .52 to .60 and for females from .46 to .54. Socialization, flexibility, achievement via conformance, and femininity were the largest contributors for both sexes.

Watson (1967)

Data were obtained from three groups of college males using the 18 subscales of the California Psychological Inventory to predict cumulative GPA. Two of the three groups were defined as maladjusted because they had sought counseling services. Very little across-group consistency was found; the only exception was intellectual efficiency (median r = .27, significant across all three groups). When percentile rank was partialed out, the correlations were no longer significant. More significant correlations were found in the maladjusted groups, but, similarly, when percentile rank was partialed out, 4 out of the 6 fell to nonsignificance. It was concluded that—with the possible exception of achievement by conformity, achievement via independence, and intellectual efficiency—the CPI scales cannot be fruitfully employed as predictors of academic success in male undergraduates.

Appendixes

Appendix A. Biographical data

Predictive validity of biographical data

Study and setting	Class or year	N	Inventory type or item	Outcome predicted	Predictive correlations r	R_1	R_2	ΔR
Aiken (1964)	1962	100M	76-item multiple choice	1st-semester				
Women's College of the University of			biographical inventory	GPA	.60			
North Carolina. A		100M	(Cross validation)		.57	.58[a]	.70[b]	.12
total of 1,006 incoming freshmen were		50F	Home 100 miles or more from college.	Freshman dropout	.05			
administered the inventory during			No relatives graduated from this college.		.05			
orientation, and two samples of 100 each			Not in top 5% of HS.		.05			
were randomly drawn for analysis. The			Not awarded scholarship.		.05			
second sample was			Not in art, creative clubs, or newspaper in HS.		.05			
used for cross-validation of item-weights developed			Sure about vocational goals.		.05			
from the first sample.			Expectation for graduating from this college not high.		.05			
Freshman women			Did not like foreign language study.		.05			
who withdrew from college before or at			Did not like television musicals.		.01			
midsemester were			Did not prefer reading for leisure.		.01			
compared with 50 randomly selected			Did not believe knows more than high school grades indicate.		.01			
nonwithdrawals.			Desire to graduate not "very strong."		.01			

a. HSR, SAT, class size

b. HSR, SAT, class size, and biographical inventory

Predictive validity of biographical data (continued)

Study and setting	Class or year	N	Inventory type or item	Outcome predicted	Predictive correlations r	R_1	R_2	ΔR
Anastasi, Meade & Schneiders (1960) Fordham University. The outcome predicted was long-term and comprehensive college performance based on extensive information from several sources within the university. All data are based on cross-validation samples.	1958, 1959	300	80-item biographical inventory	Membership in positive vs. negative criterion group	.55			
				Membership in positive vs. average criterion group	.35			
				Membership in average vs. negative criterion group	.26			
Anastasi & Schaefer (1969) 7 high schools in New York metropolitan area. Teacher's nominations used to identify creative girl students. Apparently these data are a subset of those analyzed and reported by Schaefer (1969).	Not given	400F	166-item biographical inventory	Artistic achievement	.34			
Brown & Dubois (1964) Iowa State University: high-ability freshmen. Science and humanities students	1961–1962	76	Number of hours of study per week.	First two quarters cumulative	.07			
			Thinking of classwork between classes.		.23			
			Free time between classes spent studying.		−.09			
			Student's lowest acceptable GPA.		.56			
			Parent's lowest acceptable GPA.		.46			

Predictive validity of biographical data (continued)

Study and setting	Class or year	N	Inventory type or item	Outcome predicted	Predictive correlations r	R_1	R_2	ΔR
Brown & Dubois (1964) (continued)								
Engineering students	1961–1962	125	Number of hours of study per week.		.30			
			Thinking of classwork between classes.		.10			
			Free time between classes spent studying.		−.28			
			Student's lowest acceptable GPA.		.38			
			Parent's lowest acceptable GPA.		.23			
Science and humanities students		76	Father's education, mother's education, importance of education in home.			.50[c]	.63[d]	.13
Harrington (1969) Freshman students at Ohio University were administered a 300-item biographical inventory.	1966	746M	Creativity scale	Freshman GPA	.27			
		746M	Management potential scale		.21			
		780F	Creativity scale		.30			
		780F	Management potential scale		.17			
Hilton & Myers (1976) A 169-item Background and Experience Questionnaire (BEQ) was administered to high school student in the 11th grade and used to predict senior class rank.	1961	1206	BEQ	Senior HS class rank	.57			

c. Math test score, HSA

d. Math test score, HSA, and three biographical items

Predictive validity of biographical data (continued)

Study and setting	Class or year	N	Inventory type or item	Outcome predicted	Predictive correlations r	R_1	R_2	ΔR
Korman et al. (1968) University of Texas Southwestern Medical School (Dallas). The outcomes predicted were factors based on performance until the time students were seniors in medical school	1962 (class)	68	Number of children in family.	General achievement	.04			
			Raised in large city.		.05			
			Father's education.		.01			
			Advanced grade placement.		.14			
			Earned money early.		.03			
			Frequency of teen dating.		.12			
			Many friends.		−.09			
			Class or club officer.		−.17			
			Independence encouraged.		−.02			
			Talked over problems with parents.		.13			
			Conflicts with parents.		.04			
			School behavior problem.		.24			
			Meals irregular at home.		.25			
			Open disagreements with parents.		−.23			
			Parent–child discussions.		−.07			
			Parental interest.		.06			
			Came home "too late" often.		.00			
			Neatness emphasized.		−.16			
			Sickly as adolescent.		.18			
	1962 (class)	68	Number of children in family.	Humanism (interested in public health, suitable as general practitioner; high integrity rating)	.24			
			Raised in large city.		−.36			
			Father's education.		−.09			
			Advanced grade placement.		−.28			
			Earned money early.		.01			
			Frequency of teen dating.		.10			
			Many friends.		.08			
			Class or club officer.		.03			
			Independence encouraged.		.31			
			Talked over problems with parents.		.27			
			Conflicts with parents.		−.26			
			School behavior problems.		−.15			
			Meals irregular at home.		.01			
			Open disagreements with parents.		−.23			
			Parent–child discussions.		−.03			
			Parental interest.		.04			
			Came home "too late" often.		−.23			
			Neatness emphasized.		−.06			
			Sickly as adolescent.		.02			

Predictive validity of biographical data (continued)

Study and setting	Class or year	N	Inventory type or item	Outcome predicted	Predictive correlations r	R_1	R_2	ΔR
Korman et al. (1968) (continued)	1962 (class)	68	Number of children in family.	Peer esteem (based on peer ratings of 12 variables completed by each member of graduating class)	.18			
			Raised in large city.		−.34			
			Father's education.		.06			
			Advanced grade placement.		−.02			
			Earned money early.		.07			
			Frequency of teen dating.		−.15			
			Many friends.		.15			
			Class or club officer.		.33			
			Independence encouraged.		.08			
			Talked over problems with parents.		.23			
			Conflicts with parents.		−.07			
			School behavior problem.		−.25			
			Meals irregular at home.		.02			
			Open disagreements with parents.		−.19			
			Parent–child discussions.		.25			
			Parental interest.		−.02			
			Came home "too late" often.		.10			
			Neatness emphasized.		.47			
			Sickly as adolescent.		−.14			
	1962 (class)	68	Number of children in family.	Scientist potential (suitable as researcher and teacher; interested in diagnosis vs. treatment)	−.13			
			Raised in large city.		.36			
			Father's education.		.23			
			Advanced grade placement.		.14			
			Earned money early.		−.27			
			Frequency of teen dating.		−.02			
			Many friends.		−.02			
			Class or club officer.		−.18			
			Independence encouraged.		.06			
			Talked over problems with parents.		.10			
			Conflicts with parents.		.14			
			School behavior problem.		.12			
			Meals irregular at home.		−.05			
			Open disagreements with parents.		.26			
			Parent–child discussions.		.01			
			Parental interest.		−.01			
			Came home "too late" often.		.16			
			Neatness emphasized.		−.03			
			Sickly as adolescent.		.23			

Predictive validity of biographical data (continued)

Study and setting	Class or year	N	Inventory type or item	Outcome predicted	Predictive correlations r	R_1	R_2	ΔR
Korman et al. (1968) (continued)	1962 (class)	68	Number of children in family.	Internship success (based on faculty ratings of internship performance)	−.10			
			Raised in large city.		−.06			
			Father's education.		−.03			
			Advanced grade placement.		−.20			
			Earned money early.		−.06			
			Frequency of teen dating.		.31			
			Many friends.		.27			
			Class or club officer.		.24			
			Independence encouraged.		.22			
			Talked over problems with parents.		.11			
			Conflicts with parents.		−.16			
			School behavior problems.		−.14			
			Meals irregular at home.		−.10			
			Open disagreements with parents.		−.17			
			Parent–child discussions.		−.08			
			Parental interest.		.28			
			Came home "too late" often.		−.08			
			Neatness emphasized.		.05			
			Sickly as adolescent.		−.12			
Lunneborg & Lunneborg (1966) University of Washington freshmen	1965	520	5 biographical predictors	Withdrawal from college	.35			
			8 biographical predictors	Change of major	.35			
Malloy (1955) University of Nebraska freshmen	1952	287M	Life Experience Inventory (LEI)	First semester GPA	.43	–	.53[e]	.10
		154F			.45	–	.57[e]	.12
Malloy & Ivanoff (1964) Marquette University third-year business school students (males) and second-year liberal arts students (females)	Not given	60M	Life Experience Inventory (LEI)	Sophomore cumulative GPA	.56	.57[f]	.69[g]	.12
		80F			.52	.65[f]	.70[g]	.05

e. Does not include high school grades or rank.

f. Includes high school rank, SAT-V, and SAT-M.

g. Includes high school rank, SAT-V, SAT-M, and LEI.

Predictive validity of biographical data (continued)

Study and setting	Class or year	N	Inventory type or item	Outcome predicted	Predictive correlations r	R_1	R_2	ΔR
Myers (1952) Freshmen at an Eastern women's liberal arts college	1951	355	Biographical score based on religion, extent of participation in outside activities, county of residence, whether parents native born, birth order, and number of high school offices held.	Freshman GPA	.26			
	1952	346			.29			
Myers & Schultz (1950) Freshmen at an Eastern women's liberal arts college	1952	346	A 45-item score (Score 1) based on the same questionnaire as Myers (1952)	Freshman GPA	.10			
			A 31-item score (Score 2)		.11			
			A 19-item score (Score 3)		.11			
Payne, Rapley, & Wells (1973) Freshman at a large Southern University	Not given	931M	Parental warmth	Freshman GPA	.01			
			Intellectualism		.04			
			Academic achievement		.45			
			Social introversion		.11			
			Scientific interest		−.03			
			Socioeconomic status		.08			
			Independence		.06			
			Parental control		−.15			
			Positive academic attitude		−.03			
			Sibling friction		.00			
			Religious activity		.03			
			Athletic interest		−.10			
			Social desirability		−.04			
		818F	Maternal warmth		.05			
			Social leadership		.03			
			Academic achievement		.41			
			Parental control		−.02			
			Cultural literary interests		−.02			
			Scientific-artistic interests		−.07			
			Socioeconomic status		.06			
			Expression of negative		−.01			
			Athletic participation		−.05			
			Conformity to female role		.03			
			Maladjustment		.12			
			Popularity with opposite sex		−.09			
			Positive academic attitude		.01			
			Daddy's girl		.04			
			Social maturity		−.12			
		931M	All scales combined			.40	.56	.16
		818F	All scales combined			.44	.53	.09

Predictive validity of biographical data (continued)

Study and setting	Class or year	N	Inventory type or item	Outcome predicted	Predictive correlations r	R_1	R_2	ΔR
Petry & Craft (1976) Memphis State University teacher education program.	Not given	52	ABI-academic performance	Cumulative	.45			
			ABI-creativity	GPA	.05			
			ABI-both scales		.49	.49	.67	.18
High-risk students were administered several measures, including the 300-item Alpha Biographical Inventory (ABI).		52	ABI-academic performance	Professional	.51			
			ABI-creativity	GPA (education	.13			
			ABI-both scales	courses)	.54	.49	.60	.11
Richards & Lutz (1968)	1966	3267M	Student profile section of the ACT assessment (self-reported high school achievements in specific areas).	College leadership achievement.	.23			
				College music achievement.	.32			
				College speech and drama achievement.	.21			
				College writing achievement.	.22			
				College science achievement.	.21			

Predictive validity of biographical data (continued)

Study and setting	Class or year	N	Inventory type or item	Outcome predicted	Predictive correlations r	R_1	R_2	ΔR
Richards & Lutz (1968) (continued)		2428F		College leadership achievement.	.20			
				College music achievement.	.35			
				College speech and drama achievement.	.26			
				College writing achievement.	.30			
				College science achievement.	.21			
Schaefer (1969) 10 high schools in the New York metropolitan area. Teacher's nominations for having demonstrated creativity in high school were used to identify creative students.	Not given	800	165-item biographical inventory on physical characteristics, family history, educational history, leisure-time activities, etc.	Artistic achievement (M)	.64			
				Writing achievement (F)	.55			
Schultz & Green (1953) Eastern women's liberal arts college. A continuation of the study begun by Myers & Schultz.	1953	322	Score I, based on 55 items of a 145-item questionnaire covering motivation, intellectual interests, teacher relations, and study habits.	Freshman GPA	.11			
			Score II, based on 69 items of the same questionnaire.		.16			
			Score III, based on 83 items of the same questionnaire.		.16			

Predictive validity of biographical data (continued)

Study and setting	Class or year	N	Inventory type or item	Outcome predicted	Predictive correlations r	R_1	R_2	ΔR
Scott & Castles (1975) 19 high schools of a county public school system		123M 144F	A 118-item biographical inventory	Nonacademic accomplishment in in high school	.33 .67			
Taylor & Ellison (1967) NASA scientists	1960	354	300-item biographical inventory (Form A)	Scientific creativity	.59			
Webb (1960) Emory University. Incremental validity was examined by adding the biographical inventory to high school average and two test scores.	1955 (class)	303M 172F	Emory biographical inventory, a 200-item multiple-choice inventory including items about health, socioeconomic status, family relations, and habits.	First-year college GPA (excluding physical education grades)	.41 .45	.65 .59	.66 .60	.01 .01
Willingham (1965) Georgia Tech Students. Incremental validity of biographical data was examined by adding information from an application blank to SAT-V, SAT-M, and high school average.	1960–1962	2224 2556	35-item application blank	Freshman average Voluntary withdrawal	.48 .25 (rbis)	.60	.62	.02

Appendix B. Recommendations

Predictive validity of recommendations

Study and setting	Class or year	N	Type of rating or recommendation	Outcome predicted	Predictive correlations r	R_1	R_2	ΔR
Browning (1968) Elementary school teachers (508) in the Montgomery County, Maryland, Public Schools. Four-point ratings were made by a total of 2,221 different references on each of six characteristics. The criterion (teaching competency) was based on one full year (9 months) of observation.	Not given	508 (teachers)	Teaching ability	Overall rating of teaching competency	.03			
			Relationships with children		.08			
			Community relationships		.06			
			Professional relationships		.04			
			Personal qualities		.06			
			Health		.02			
		147	Combined ratings	Last superintendents' ratings	.08			
		340		Last principals' ratings	.19			
		96		Last supervisors' ratings	.23			
		111		Other superintendents' ratings	.22			
		407		Other principals' ratings	.07			
		53		Other supervisors' ratings	.06			
		150		Heads of departments' ratings	.21			
		68		Practice teaching professors' ratings	−.03			
		184		Practice teaching teachers' ratings	.09			
		32		Deans of college ratings	.19			
		633		Other ratings	.18			
		2,221 (ratings)	Combined totals	Overall rating of teaching competency	.13			

116

Predictive validity of recommendations (continued)

Study and setting	Class or year	N	Type of rating or recommendation	Outcome predicted	Predictive correlations			
					r	R_1	R_2	ΔR
Calkins et al. (1974) University of Missouri (Kansas City) School of Medicine	1972 (appl.)	162	Reference ratings	Selection Council Index (used for selection decisions)	.38			
	1973 (appl.)	241			.38			
	1974 (appl.)	335			.28			
Calkins et al. (1980) University of Missouri (Kansas City) School of Medicine	1972–1977 entering classes	396 non minorities	Reference ratings	Two-year cumulative GPA	.24			
				Second-year examination	.24			
				Clinical performance evaluation	.16			
				Status in program at end of second year	.16			
	1972–1977 entering classes	51 minorities	Reference ratings	Two-year cumulative GPA	.29			
				Second-year examination	.11			
				Clinical performance evaluation	.34			
				Status in program at end of second year	.07			
Creager (1965) National Science Foundation Fellowship applicants	1955–1956	520M	Reference ratings	Ph.D. attainment	.12	.25	.26	.01
		191F			.08	.35	.36	.01
		663M			.18	.43	.43	.00
		84F			.13	.55	.55	.00
		585M			.17	.27	.27	.00
		176			.21	.47	.47	.00
		264M			.20	.41	.41	.00
		769M			.15	.39	.39	.00
		161			.23	.42	.42	.00

Predictive validity of recommendations (continued)

Study and setting	Class or year	N	Type of rating or recommendation	Outcome predicted	Predictive correlations r	R_1	R_2	ΔR
Creager (1965) (continued)	1955–1956	520M		B.A.-to-Ph.D. time lapse	.23	.37	.38	.01
		191F			.13	.30	.30	.00
		663M			.24	.41	.41	.00
		84F			.11	.50	.50	.00
		585M			.22	.30	.31	.01
		176			.33	.40	.46	.06
		264M			.27	.42	.44	.02
		769M			.23	.42	.42	.00
		161M			.23	.32	.35	.03
Creager & Harmon (1966)	1955	182	Biology applicant reference ratings	Scientific knowledge rating several years after application	.29			
	1956	160						
Candidates for NSF predoctoral fellowships were followed up several years later and various measures of performance obtained.	1955	185	Chemistry applicant reference ratings		.36			
	1956	171			.38			
					.36			
	1955	249	Engineering applicant reference ratings					
	1956				.30			
		109	Geology applicant reference ratings		.34			
		134	Mathematics applicant reference ratings		.08			
	1955	192	Physics applicant reference ratings		.26			
	1956	160			.25			
	1955, 1956	73	Psychology applicant reference ratings		.11			
	1955	182	Biology applicant reference ratings	Overall ability rating several years after application	.22			
	1956	160			.33			
	1955	185	Chemistry applicant reference ratings		.34			
	1956	171			.35			
	1955, 1956	249	Engineering applicant reference ratings		.28			
		109	Geology applicant reference ratings		.31			
		134	Mathematics applicant reference ratings		.16			
	1955	192	Physics applicant reference ratings		.27			
	1956	160			.27			
	1955, 1956	73	Psychology applicant reference ratings		.13			

Predictive validity of recommendations (continued)

Study and setting	Class or year	N	Type of rating or recommendation	Outcome predicted	Predictive correlations r	R_1	R_2	ΔR
Mehrabian (1969) Graduate psychology student selection at UCLA.	Not given	79	Overall promise as a graduate psychology student (made by graduate psychology professors)	Overall graduate school performance index	.36	.43[a]	.51	.08
				Professor's evalution of graduate achievement	.25			
				Graduate grades in content courses	.25			
				Graduate grades in statistics courses	.38			
Mosell & Goheen (1958) Supervisors' ratings in twelve skilled occupations were compared with the Employment Recommendation Questionnaire (ERQ) obtained in the standard federal application for employment	Not given	51	Carpenter	On-the-job supervisors' ratings	.01			
		40	Equipment repairmen		.23			
		100	Machinist		.24			
		108	Machine operator		−.10			
		125	Ordnanceman – torpedo		−.01			
		107	Radio mechanic		.29			
		94	Aviation metalsmith		.24			
		108	Highlift fork operator		.21			
		98	Auto mechanic		.09			
		70	Painter		.07			
		100	Ordnanceman		.10			
		116	Printer		.11			
Mosell & Goheen (1959) Applicants for federal service.	Not given	102	Personnel officers	Supervisors' ratings	.02			
		188	Previous supervisors		.19			
		311	Previous co-workers		.09			
		182	Acquaintances		.20			
		12	Relatives		−.16			

a. Note that this baseline R is based on the combination of GRE and MCAT scores only; undergraduate GPA is not included.

Predictive validity of recommendations (continued)

Study and setting	Class or year	N	Type of rating or recommendation	Outcome predicted	Predictive correlations			
					r	R_1	R_2	ΔR
Mosell & Goheen (1959) (continued)		51 (carpenters)	Combined personnel officers and previous supervisors	Supervisors' ratings	.05			
		40 (equipment repairmen)			.38			
		76 (machinists)			−.09			
		108 (machine operators)			.42			
		125 (ordnancemen)			−.20			
Nicholson (1970) Brown University. This was a long-term study of several predictors and several criteria. Correlations were reported only for the prediction of freshman GPA, but predictions were also made of graduation, attainment of advanced degree, and a long-term (15–17 years) alumni rating of success.	1950–1952	1,222 nonveterans	A composite of several ratings made by secondary school counselors.	Freshman GPA	.28			
	1950–1952	905 veterans			.22			
	1966–1968	1,616			.30			
	1970	583			.37			
	1971	689			.29			

Predictive validity of recommendations (continued)

Study and setting	Class or year	N	Type of rating or recommendation	Outcome predicted	Predictive correlations r	R_1	R_2	ΔR
Reilly (1971) Law school admission at one Far West school. One standard recommendation form was available for each student.	Not given	152	Native intelligence	First-year grades	.30			<.01
			Independence of thought		.16			<.01
			Oral communication		.23	.56[b]	.58	.02
			Written communication		.27			<.01
			Industry and motivation		.22			<.01
			Judgment and maturity		.26			<.01
			Leadership ability		.11			<.01
			Personality		.13			<.01
		76	Oral communication		—	.68[c]	.71	.03
Richards & Taylor (1961) University of Utah College of Medicine	1955, 1956, 1957 (admissions years)	57	Premedical advisory committee ratings	First-year medical school average	.25			
				Second-year medical school average	.31	.42[d]	.45	.03
				Third-year medical school average	.03			

b. Based on a combination of undergraduate GPA and LSAT score.

c. Based on a combination of undergraduate GPA and a writing ability test.

d. R_1 is based on undergraduate average only.

Appendix C. Admissions Interviews

Predictive validity of admissions interviews

Study and setting	Class or year	N	Interview type or rating made	Outcome predicted	Predictive correlations r
Asher (1970b)		20			
Statistics course			Structured, 10 independent raters	Performance in statistics course	.44
			Unstructured, 8 independent raters		.38
Calkins, Willoughby, & Arnold, (1980)	1972–1977	396	Nonphysician interview rating	Two-year cumulative GPA	.10
University of Missouri (Kansas City) School of Medicine (nonminority applicants). The medical school program studied is a six-year program. Thus, the two years represented here are primarily the first two years of a liberal arts program.			Nonphysician interview recommendation	Two-year cumulative GPA	.06
			Nonphysician interview rating	Second-year examination	−.09
			Nonphysician interview recommendation	Second-year examination	.11
			Nonphysician interview rating	Clinical performance evaluation	.00
			Nonphysician interview recommendation	Clinical performance evaluation	.15
			Nonphysician interview rating	Status in program at end of second year	.09
			Nonphysician interview recommendation	Status in program at end of second year	.10
			Physician interview rating	Two-year cumulative GPA	.12
			Physician interview recommendation	Two-year cumulative GPA	−.02
			Physician interview rating	Second-year examination	−.01
			Physician interview recommendation	Second-year examination	.13
			Physician interview rating	Clinical performance evaluation	.05
			Physician interview recommendation	Clinical performance evaluation	.03
			Physician interview rating	Status in program at end of second term	.09
			Physician interview recommendation	Status in program at end of second term	−.10

Predictive validity of admissions interviews (continued)

Study and setting	Class or year	N	Interview type or rating made	Outcome predicted	Predictive correlations r
Calkins, Willoughby, & Arnold (1980) (continued)	1972–1977	51	Nonphysician interview rating	Two-year cumulative GPA	.32
			Nonphysician interview recommendation	Two-year cumulative GPA	−.02
			Nonphysician interview rating	Second-year examination	−.03
			Nonphysician interview recommendation	Second-year examination	.23
			Nonphysician interview rating	Clinical performance evaluation	.34
			Nonphysician interview recommendation	Clinical performance evaluation	−.05
			Nonphysician interview rating	Status in program at end of second year	.22
			Nonphysician interview recommendation	Status in program at end of second year	.02
Calkins, Willoughby, & Arnold (1980) University of Missouri (Kansas City) School of Medicine (minority applicants)	1972–1977	51	Physician interview rating	Two-year cumulative GPA	.19
			Physician interview recommendation	Two-year cumulative GPA	.15
			Physician interview rating	Second-year examination	−.03
			Physician interview recommendation	Second-year examination	.18
			Physician interview rating	Clinical performance evaluation	.25
			Physician interview recommendation	Clinical performance evaluation	.25
			Physician interview rating	Status in program at end of second year	.07
			Physician interview recommendation	Status in program at end of second year	.04
Gough & Hall (1975) University of California School of Medicine (San Francisco)	1960–1963	331 (odd)	Mean of two or more admissions interview ratings	Clinical performance factor	.09
	1966–1967	330 (even)			.16
		331 (odd)		Academic performance factor	.03
		330 (even)			.06

Predictive validity of admissions interviews (continued)

Study and setting	Class or year	N	Interview type or rating made	Outcome predicted	Predictive correlations r
Gough, Hall, & Harris (1963)	1951	70	Unspecified	First-year GPA	−.13
University of California	1952	64			−.04
School of Medicine	1953	68			−.06
(San Francisco)	1954	66			−.07
	1955	71			.15
	1956	78			.14
	1957	74			.10
	1958	83			−.09
	1959	79			−.02
	1960	78			.10
	1961	70			.01
	1962	94			.05
	1963	93			.22
	1964	100			.01
	1951	70		Second-year GPA	
	1952	64			
	1953	68			.01
	1954	66			−.04
	1955	71			.06
	1956	78			.05
	1957	74			.14
	1958	83			−.19
	1959	79			.04
	1960	78			.06
	1961	70			.04
	1962	94			−.10
	1963	93			.35
	1964	100			
	1951	70		Third-year GPA	
	1952	64			.16
	1953	68			−.08
	1954	66			.05
	1955	71			.20
	1956	78			−.02
	1957	74			.32
	1958	83			.03
	1959	79			.17
	1960	78			.05
	1961	70			.02
	1962	94			.02
	1963	93			
	1964	100			

Predictive validity of admissions interviews (continued)

Study and setting	Class or year	N	Interview type or rating made	Outcome predicted	Predictive correlations r
Gough, Hall, & Harris (1963) (continued)	1951	70	Unspecified	Fourth-year GPA	−.01
	1952	64			.05
	1953	68			−.07
	1954	66			.20
	1955	71			.13
	1956	78			−.04
	1957	74			.15
	1958	83			.00
	1959	79			.14
	1960	78			−.04
	1961	70			.06
	1962	94			–
	1963	93			–
	1964	100			–
Gough, Hall, & Harris (1964) University of California School of Medicine (San Francisco)	1958	79	Unspecified	Faculty rating factor I	.07
				Faculty rating factor II	.10
				Faculty rating factor III	.24
	1959	79	Unspecified	Faculty rating factor I	.14
				Faculty rating factor II	.19
	1959	81	Admissions committee interviewers' ratings	Peer nominations	
				Cluster I	−.12
				Cluster II	−.02
				Cluster III	−.03
				A priori scoring	−.11
Keck, Arnold, Willoughby, & Calkins (1979) University of Missouri (Kansas City) School of Medicine Physician Residency Training	1971–1978	55	Physician interview	Comprehensive residency performance evaluation	.19
Kelley (1957)		112	Unspecified	Five graduation criteria	.05

Predictive validity of admissions interviews (continued)

Study and setting	Class or year	N	Interview type or rating made	Outcome predicted	Predictive correlations r
Korman et al. (1968) University of Texas Southwestern Medical School (Dallas)	1962	68	Social awareness	General achievement factor based on GPA primarily	−.10
			Intellectual curiosity		.18
			Broad interests		.05
			Theoretically oriented		.09
			Poise		.30
			Intelligence		.01
			Industriousness		.00
			Prestige and prominence		.00
			Acceptability as a student		.22
			Suitable as G.P.		−.09
			Do well in research		.09
			Social awareness	Internship success factor	.26
			Intellectual curiosity		.12
			Broad interests		.01
			Theoretically oriented		−.13
			Poise		.05
			Intelligence		−.03
			Industriousness		.11
			Prestige and prominence		.05
			Acceptability as a student		.13
			Suitable as G.P.		.11
			Do well in research		.04
			Social awareness	Peer esteem (suitable as partner, skilled in patient relationship)	.16
			Intellectual curiosity		.07
			Broad interests		.01
			Theoretically oriented		.01
			Poise		.25
			Intelligence		.10
			Industriousness		−.02
			Prestige and prominence		.17
			Acceptability as student		.17
			Suitable as G.P.		−.13
			Do well in research		−.13

Predictive validity of admissions interviews (continued)

Study and setting	Class or year	N	Interview type or rating made	Outcome predicted	Predictive correlations r
Korman et al. (1968) (continued)	1962	68	Social awareness	Scientist potential	.10
			Intellectual curiosity	(suitable as researcher	.37
			Broad interests	and teacher, interested	.23
			Theoretically oriented	in diagnosis versus	.39
			Poise	treatment)	.00
			Intelligence		.33
			Industriousness		.24
			Prestige and prominence		.12
			Acceptability as student		−.01
			Suitable as G.P.		−.03
			Do well in research		.24
			Social awareness	Humanism (interested in	−.02
			Intellectual curiosity	public health, suitable	.03
			Broad interests	as general practitioner,	.00
			Theoretically oriented	high internship rating	−.18
			Poise	on integrity)	−.04
			Intelligence		−.12
			Industriousness		−.05
			Prestige and prominence		−.26
			Acceptable as student		.07
			Suitable as G.P.		−.08
			Do well in research		.07
Linn & Winograd (1969)	1965–	332	Unspecified	First-year GPA	.00
New York University	1967	177		Second-year GPA	−.05
Law School		83		Third-year GPA	−.02
		81		Cumulative GPA	−.03
Palacios et al. (1966)		144	Highly structured, seven scales developed and combined for prediction	Job success	.48
Occupational achievement of blind adults					
Richards & Taylor (1961)	1951–	65	Group rated applicants	First-year GPA	.16
University of Utah	1957		on "personal	Second-year GPA	.23
College of Medicine			characteristics"	Third-year GPA	.22
	1951–				
	1957		Group rated applicants	First-year GPA	.23
			on "chances of	Second-year GPA	.23
			success"	Third-year GPA	.15
	1951–		Group rated applicants	First-year GPA	.22
	1957		on "recommended	Second-year GPA	.16
			decisions"	Third-year GPA	.32

Appendix D. Interest measures

Predictive validity of interest measures

Study and setting	Class or year	N	Interest measure	Outcome predicted	Predictive correlations r	R_1	R_2	ΔR
Ayers & Rohr (1974)	1967–1972	400	Kuder Preference Record −1	GPA in a teacher-preparation program (secondary school)	−.24			
			−6		−.05			
			−9		−.24			
					−.05			
					.22			
		203F	Kuder −1	GPA in a teacher-preparation program (elementary school)	−.13			
			−6		−.13			
			−9		−.06			
Campbell (1966)		250M	SVIB Achievement Scale	Freshman GPA	.36			
Harrington (1969) Freshman students at Ohio University	1966	746M	SVIB Achievement Scale	Freshman GPA	.29	.52[a]	.53[b]	.01
Keck et al. (1979) University of Missouri: Kansas City School of Medicine (a six-year combined program). The interest measure was the Self-Directed Search (SDS).	1978	56	SDS-Realistic	Medical resident evaluation by supervisors (33 seven-point scale)	.13			
			-Investigative		.02			
			-Artistic		.10			
			-Social		.06			
			-Enterprising		.05			
			-Conventional		−.23			

a. Using ACT battery and HS rank

b. Using ACT battery, HS rank, and SVIB Achievement Scale

Predictive validity of interest measures (continued)

Study and Setting	Class or Year	N	Interest measure	Outcome predicted	Predictive Correlations r	R_1	R_2	ΔR
Kegel-Flom (1975) University of California School of Medicine: San Francisco. The entire graduating class in residence in 54 medical facilities was rated on internship performance by supervisors, peers, and themselves. Those ratings were then related to pre-admission and medical school variables.	1968 Class	110	SVIB-Physician	Internship performance Supervisors' rating	.25			
			-Production manager		−.23			
			-Office worker		−.22			
			-Credit mgr.		−.22			
			-Business teacher		−.22			
			-Chamber of commerce exc.		−.20			
			-Life ins. salesman		−.19			
			-Banker		−.07			
			-Mortician		−.06			
			-Librarian		.13			
			-Purchasing agent		−.10			
			-Sales mgr.		−.16			
			-Real estate sales		−.08			
			-Pharmacist		−.02			
			SVIB-Physician	Internship performance Peer rating	−.07			
			-Production manager		−.13			
			-Office worker		−.11			
			-Credit mgr.		−.16			
			-Business teacher		−.04			
			-Chamber of commerce exec.		−.17			
			-Life ins. salesman		−.35			
			-Banker		−.14			
			-Mortician		−.21			
			-Librarian		.34			
			-Purchasing agent		−.28			
			-Sales mgr.		−.35			
			-Real estate sales		−.27			
			-Pharmacist		−.33			

Predictive validity of interest measures (continued)

Study and Setting	Class or Year	N	Interest measure	Outcome predicted	Predictive Correlations r	R_1	R_2	ΔR
Linday & Althouse (1969)	1966	299M	SVIB Achievement	Freshman GPA	.10	.41	.42	.01
		89F			.25	.52	.53	.01
Both the SVIB and the CSQ were administered to random samples of freshmen at Pennsylvania State University.		299M	CSQ-1 (motivation for grades)		.27	.41	.45	.04
		89F			.29	.52	.55	.03
Nichols (1966)	1962 (cross-validation sample)	179M	VPI grade scale	Freshman grades	.11			
		138F			.12			
Several measures, including Holland's Vocational Preference Inventory (VPI), were used to predict freshman grades and extracurricular achievement in both normative and a cross-validation sample of National Merit Scholarship participants.		179M	VPI extracurricular scale	Extracurricular achievement	.18			
		138F			−.06			
	1962 (normative sample)	201M	VPI grade scale	Freshman grades	.10			
		218F			.13			
		201M	VPI extracurricular scale	Extracurricular achievement	.22			
		218F			.07			

Predictive validity of interest measures (continued)

Study and setting	Class or year	N	Interest measure	Outcome predicted	Predictive correlations r	R_1	R_2	ΔR
Nichols & Holland (1963)	1960	544M	*Preferences*					
National Merit finalists	HS grads		Realistic	Leadership during the first year of college	−.20			
			Intellectual		ns			
			Social		.21			
			Conventional		ns			
			Enterprising		ns			
			Artistic		ns			
A comprehensive set of predictors related to a number of academic criteria of college performance. For an interest measure, students were asked to rank their preference for six groups of occupations based on Holland's VPI groups.		275F	Realistic		−.13			
			Intellectual		ns			
			Social		.14			
			Conventional		ns			
			Enterprising		ns			
			Artistic		ns			
		544M	Realistic	Literary achivement during the first year of college	−.37			
			Intellectual		ns			
			Social		ns			
			Conventional		−.16			
			Enterprising		ns			
			Artistic		.40			
		275F	Realistic		−.08			
			Intellectual		ns			
			Social		ns			
			Conventional		−.26			
			Enterprising		ns			
			Artistic		.38			
		544M	Realistic	Scientific achievement during the first year of college	−.06			
			Intellectual		ns			
			Social		ns			
			Conventional		ns			
			Enterprising		ns			
			Artistic		ns			
		275F	Realistic		.43			
			Intellectual		ns			
			Social		ns			
			Conventional		ns			
			Enterprising		ns			
			Artistic		ns			
		544M	Realistic	Achievement in graphic arts during the first year of college	ns			
			Intellectual		ns			
			Social		ns			
			Conventional		.07			
			Enterprising		ns			
			Artistic		.24			

Predictive validity of interest measures (continued)

Study and setting	Class or year	N	Interest measure	Outcome predicted		Predictive correlations r	R_1[c]	R_2[d]	ΔR
Norris & Katz (1970) Academic Interest Measures (AIM) and Student Interest (SI) ratings were studied as predictors of freshman course grades (self-reported) and interests in a pooled national sample of students in different colleges.	1969	300–350	SI (Grade 12)	Biological science interests	(M)	.52	.09	.52	.43
					(F)	.48	.07	.48	.41
				Physical science interests	(M)	.42	.13	.42	.29
					(F)	.34	.14	.38	.24
				Mathematics interests	(M)	.58	.28	.59	.31
					(F)	.64	.44	.66	.22
				Social science interests	(M)	.40	.22	.44	.22
					(F)	.30	.17	.34	.17
				Humanities interests	(M)	.40	.25	.44	.19
					(F)	.39	.25	.41	.19
				Art/arch. interests	(M)	.49	.08	.49	.41
					(F)	.47	.12	.48	.36
				Education interests	(M)	.14	.12	.16	.04
					(F)	.10	.16	.18	.02
				Business interests	(M)	.50	.23	.51	.28
					(F)	.58	.33	.60	.27
				Engineering interests	(M)	.31	.17	.38	.21
					(F)	.23	.15	.29	.14
				Home econ., etc. interests	(M)[e]	.17	.25	.29	.04
					(F)	.39	.25	.42	.17
				Applied science interests	(M)	.36	.07	.37	.30
					(F)	.21	.15	.24	.09
			AIM (Grade 11)	Biological science interests	(M)	.37	.09	.38	.29
					(F)	.34	.07	.35	.28
				Physical science interests	(M)	.32	.13	.34	.21
					(F)	.32	.14	.34	.20
				Mathematics interests	(M)	.42	.28	.45	.17
					(F)	.57	.44	.60	.16
				Social science interests	(M)	.38	.22	.40	.18
					(F)	.22	.17	.26	.09
				Humanities interests	(M)	.41	.25	.43	.18
					(F)	.33	.25	.36	.11
				Art/arch. interests	(M)	.46	.08	.46	.38
					(F)	.43	.12	.43	.31
				Education interests	(M)	.17	.12	.18	.06
					(F)	.02	.16	.17	.01
				Business interests	(M)	.35	.23	.39	.16
					(F)	.40	.33	.48	.15

c. Using PSAT-V and PSAT-M

d. Using PSAT-V, PSAT-M, and AIM

e. Includes also health and physical education

Predictive validity of interest measures (continued)

Study and setting	Class or year	N	Interest measure	Outcome predicted		Predictive correlations r	R_1[c]	R_2[d]	ΔR
Norris & Katz (1970) (continued)	1969	300–350	AIM (Grade 11)	Engineering interests	(M)	.30	.17	.44	.27
					(F)	.26	.15	.32	.17
				Home econ., etc. interests[e]	(M)	.14	.25	.27	.02
					(F)	.24	.25	.31	.06
				Applied science interests	(M)	.37	.07	.37	.30
					(F)	.29	.15	.31	.16
		492M	AIM	Biological science grades		.14	.34	.35	.01
		844F	(Grade 11)	Biological science grades		.14	.31	.33	.02
		755M		Physical science grades		.09	.22	.23	.01
		603F		Physical science grades		.12	.27	.28	.01
		1,209M		Mathematics grades		.13	.24	.25	.01
		1,050F		Mathematics grades		.22	.33	.34	.01
		1,035M		Social science grades		.24	.31	.36	.05
		1,494F		Social science grades		.14	.30	.31	.01
		1,193M		Humanities grades		.16	.33	.34	.01
		1,573F		Humanities grades		.14	.27	.28	.01
		272M		Art/arch. grades		.15	.14	.20	.06
		560F		Art/arch. grades		.06	.19	.19	.00
		123M		Education grades		.13	.14	.17	.03
		227F		Education grades		.08	.19	.20	.01
		303M		Business grades		−.06	.35	.37	.02
		306M		Business grades		.02	.19	.20	.01
		260M		Engineering grades		.11	.35	.35	.00
		1,350F		Home econ., etc. grades		.02	.09	.09	.00
		280M		Applied science grades		.06	.25	.26	.01
		104F		Applied science grades		.09	.11	.12	.01

c. Using PSAT-V and PSAT-M

d. Using PSAT-V, PSAT-M, and AIM

e. Includes also health and physical education

Predictive validity of interest measures (continued)

Study and Setting	Class or Year	N	Interest measure	Outcome predicted	Predictive Correlations r	R_1	R_2	ΔR
Norris & Katz (1970) (continued)	1969	492M	AIM plus SI	Biological science grades	–	–	.36	–
		844F		Biological science grades	–	–	.34	–
		755M		Physical science grades	–	–	.23	–
		603F		Physical science grades	–	–	.28	–
		1,209M		Mathematics grades	–	.31	.31	.00
		1,050F		Mathematics grades	–	.42	.42	.00
		1,035M		Social science grades	–	.40	.42	.02
		1,494F		Social science grades	–	.36	.36	.00
		1,193M		Humanities grades	–	.39	.41	.02
		1,573F		Humanities grades	–	.31	.34	.03
		272M		Art/arch. grades	–	–	.30	–
		560F		Art/arch. grades	–	.18	.22	.04
		123M		Education grades	–	.24	.32	.08
		227F		Education grades	–	–	.26	–
		303M		Business grades	–	–	.37	–
		306M		Business grades	–	–	.32	–
		260M		Engineering grades	–	–	.38	–
		1,350F		Home econ., etc. grades[e]	–	–	.10	–
		280M		Applied science grades	–	–	.26	–
		104F		Applied science grades	–	–	.12	–
		492M	SI	Biological science grades	.15	.34	.35	.01
		844F	(grade 12)	Biological science grades	.12	.31	.33	.02
		755M		Physical science grades	.09	.22	.22	.00
		603F		Physical science grades	.09	.22	.22	.00
		1,209M		Mathematics grades	.09	.27	.28	.01
		1,050F		Mathematics grades	.11	.24	.25	.01
		1,035M		Social science grades	.27	.33	.27	.04
		1,494F		Social science grades	.16	.31	.34	.03
		1,193M		Humanities grades	.11	.30	.31	.01
		1,573F		Humanities grades	.20	.33	.35	.02
		272M		Art/arch. grades	.18	.27	.29	.02
		560F		Art/arch. grades	.27	.14	.29	.15
		123M		Education grades	.12	.19	.22	.03
		227F		Education grades	−.09	.14	.19	.05
		303M		Business grades	.09	.19	.21	.02
		306M		Business grades	.02	.35	.37	.02
		260M		Engineering grades	.08	.19	.28	.09
		1,350F		Home econ., etc. grades[e]	.03	.09	.10	.01
		280M		Applied science grades	.01	.25	.25	.00
		104F		Applied science grades	.06	.11	.11	.00

c. Using PSAT-V and PSAT-M as predictors

d. Using PSAT-V, PSAT-M, and SI as predictors

e. Includes also health and physical education

Appendix E. Personality measures

Predictive validity of personality measures

Study and setting	Class or year	N	Personality measure	Outcome predicted	Predictive correlations r	R_1	R_2	ΔR
Ayers & Rohr (1974) Teacher preparation program at Tennessee Technological University	1967–1972	400	SSHA	GPA in a teacher preparation program (secondary school)	−.55			
		203 (F)		GPA in a teacher preparation program (elementary school)	.12			
Brown & Dubois (1964)	1961–1962	105 science and humanities students	MMPI-Pd	Freshman GPA	−.16			
			-Mf		.28			
			-Ma		−.20			
			-Sie		.22			
			-As		.23			
			-composite			.63	.65	.02
Donnan (1968) University of North Carolina	1962	375	OAIS (achiever personality)	Freshman GPA	.41	.56	.63	.07
				Sophomore cum. GPA	.41	.55	.62	.07
				First-year survival	.23	.29	.34	.05
				Two-year survival	.25	.33	.38	.05
	1964 (all freshmen)	1747		Freshman GPA	.36	.55	.59	.04
				First-year survival	.13	.21	.22	.01
Dunham (1973) Hanover College (Indiana). Freshmen were administered a battery of personality tests.	1970	303	EPPS-n Ach.	Freshman GPA	.11			
		78	TAT-n Ach.		.22			
		303	Abasement/assurance (Stern).		−.01			
		303	Activities index—n ach.		−.01			
			Activities index—ego ach.		.17			
			Activities index—fantasied—ach.		.02			

Predictive validity of personality measures (continued)

Study and setting	Class or year	N	Personality measure	Outcome predicted	Predictive correlations r	R_1	R_2	ΔR
Gough & Hall (1975) University of California School of Medicine (San Francisco)	1960–1963, 1966–1967	331 (even) 330 (odd)	ACL	Clinical performance factor	.13 .11			
		331 (even)	ACL	Academic performance factor	.06	.44		
		330 (odd)			.08	.43		
		661	ACL	Clinical performance factor	.25	.46	.47	.01
Himelstein (1965) New Mexico State University. Introducing psychology students	Not given	281	MMPI-Ac	College GPA	.37			
			-Ae		.16			
			-Gr		.28			
			-Hr		.39			
			-Ie		.34			
			-Or		.24			
			-Un		−.03			
		193	MMPI-Ac	ACT score	.40			
			-Ae		.22			
			-Gr		.30			
			-Hr		.44			
			-Ie		.37			
			-Or		.24			
			-Un		−.14			
Keck et al. (1978) University of Missouri (Kansas City) School of Medicine. Physicians who had graduated prior to 1978 from the six-year combined B.A./M.D. program participated in a follow-up study when they were resident interns.	1978 residents	56	OPI-Autonomy	Resident Evaluation Form containing 33 seven-point scales of performance in these areas: professional responsibility, relations with colleagues, self-appraisal ability, clinical skills, relation with patients, work in groups, critical thinking ability, independent learning ability, and problem solving.	.24			
			-Complexity		−.19			
			-Theoretical		.13			

Predictive validity of personality measures (continued)

Study and setting	Class or year	N	Personality measure	Outcome predicted	Predictive correlations r	R_1	R_2	ΔR
Kegel-Flom (1975) University of California School of Medicine, San Francisco. A total of 110 interns in residence at 54 different medical schools were rated at the end of their internship year by both supervisors and peers.	1968 grads.	110	CPI-Dominance	Supervisor rating at end of internship year	.01			
			-Flexibility		−.03			
			-Tolerance		.08			
			CPI-Dominance	Peer rating at end of internship year	.30			
			-Flexibility		−.23			
			-Tolerance		−.03			
Korman et al. (1968) University of Texas Southwestern Medical School (Dallas)	1962 (class)	68	CPI-	Scientist potential (suitable as researcher and teacher, interested in diagnosis vs. treatment)				
			-capacity for status		.15			
			-responsibility		.09			
			-self-control		.18			
			-tolerance		−.17			
			-good impression		.09			
			-ach. by conformity		.26			
			-ach. by independence		.33			
			-psychological minded		.30			
			-femininity		.18			
			CPI	Humanism (interested in public health, suitable as G.P., internship rating of high integrity)				
			-capacity for status		−.07			
			-responsibility		−.07			
			-self-control		−.07			
			-tolerance		−.04			
			-good impression		−.17			
			-ach. by conformity		−.02			
			-ach. by independence		−.08			
			-psychological minded		−.03			
			-femininity		−.01			
			Gough Medical School Prediction Index	General ach. factor	.00			
				Internship success	.08			
				Peer esteem	.02			
				Scientist potential	−.06			
				Humanism	−.01			

Predictive validity of personality measures (continued)

Study and setting	Class or year	N	Personality measure	Outcome predicted	Predictive correlations r	R_1	R_2	ΔR
Korman et al. (1968) (continued)	1962 (class)	68	CPI	General ach. factor				
			-capacity for status		.17			
			-responsibility		.08			
			-self-control		.05			
			-tolerance		.11			
			-good impression		−.03			
			-ach. by conformity		.09			
			-ach. by independence		.13			
			-psychological minded		.15			
			-femininity		.04			
			CPI	Internship success				
			-capacity for status		−.01			
			-responsibility		−.01			
			-self-control		−.07			
			-tolerance		−.07			
			-good impression		−.06			
			-ach. by conformity		.11			
			-ach. by independence		.06			
			-psychological minded		−.06			
			-femininity		−.06			
			CPI	Peer esteem				
			-capacity for status		−.24			
			-responsibility		−.27			
			-self-control		−.28			
			-tolerance		−.23			
			-good impression		−.23			
			-ach. by conformity		−.15			
			-ach. by independence		−.24			
			-psychological minded		−.31			
			-femininity		−.25			
Nichols & Holland (1963) National Merit finalists were administered an extensive set of measures and then followed up as college freshmen. Except for the prediction of grades, all correlations reported are biserials.	1960	544M 275F	CPI socialization scale	Freshman grades	.20 .14			
			CPI superego scale		.10 .16			
			16 PF super-ego strength scale		.20 .12			
			Self-rating: dependability		.22 .11			

Predictive validity of personality measures (continued)

Study and Setting	Class or Year	N	Personality measure	Outcome predicted	Predictive Correlations r	R_1	R_2	ΔR
Nichols & Holland (1963) (continued)	1960	544M 275F	Extraversion/introversion	Leadership achievement	—			
			Self-rating: popularity		.15			
			Self-rating: sociability		.15			
			Extraversion/introversion		.20			
			Sense of destiny scale (Gough)		.13			
					.11			
					.23			
					.32			
					.07			
			Extraversion/introversion	Dramatic achievement	.26			
					.30			
			16 PF-dominance scale		.24			
					.13			
			CPI social presence		.22			
					.33			
			Complexity/simplicity scale (Barron)	Literary achievement	.11			
					.24			
			Intuition/sensation (Myers-Briggs)		.20			
					.24			
			Perception/judging (Myers-Briggs)		.10			
					.20			
			Feeling/thinking (Myers-Briggs)		.19			
					−.08			
			CPI socialization scale (reversed)		.00			
					.20			
			16 PF dominance scale		.20			
					.26			
			Sense of destiny scale (Gough)		.19			
					.23			
			16 PF dominance scale	Rare literary achievement	.09			
					.30			
			Sense of destiny scale (Gough)		.23			
					.32			

Predictive validity of personality measures (continued)

Study and setting	Class or year	N	Personality measure	Outcome predicted	Predictive correlations r	R_1	R_2	ΔR
Nichols & Holland (1963), (continued)	1960	544M 275F	Tolerance of ambiguity	Ph.D. aspiration	.18 .25			
			Complexity/simplicity		.14 .35			
			Intuition/sensation		.14 .37			
			Feeling/thinking		.05 −.21			
			CPI femininity		.24 −.06			
			Radicalism (16 PF)		.19 .14			
Petry & Craft (1976) Memphis State University. High-risk students in a teacher training program were administered the Allport-Vernon Study of Values (SOV) and the Thorndike Dimensions of Temperament (TDOT).	Not given	52	SOV-theoretical	Cumulative GPA	.18			
			-commercial		−.22			
			-aesthetic		−.33			
			-social		.13			
			-political		.13			
			-religious		−.13			
			SOV-all scales		.44	.49[a]	.60[b]	.11
			SOV-theoretical	Professional GPA	.13			
			-economical	(education courses)	−.09			
			-aesthetic		−.28			
			-social		.27			
			-political		.02			
			-religion		−.16			
			SOV-all scales		.39	.49[a]	.59[b]	.10
			TDOT-all scales	Cumulative GPA	.52	.49[a]	.67[b]	.18
			TDOT-all scales	Professional GPA	.28	.49[a]	.57[b]	.08

a. Using SCAT-V and SCAT-M
b. Using SCAT-V, SCAT-M, and personality measure

Predictive validity of personality measures (continued)

Study and setting	Class or year	N	Personality measure	Outcome predicted	Predictive correlations r[c]	R_1	R_2	ΔR
Watson (1967) University of Iowa freshmen, freshmen who sought counseling, and upperclassmen who sought counseling.	1957–1960	100	CPI-dominance	GPA	.17			
			-cap. for status		.18			
			-sociability		.18			
			-social presence		.08			
			-self-acceptance		.12			
			-well-being		.07			
			-responsibility		.19			
			-socialization		.09			
			-self-control		.00			
			-tolerance		.21			
			-good impression		.00			
			-communality		−.01			
			-ach. by conformity		.16			
			-ach. by independence		.38			
			-intellectual efficiency		.27			
			-psychological minded		.18			
			-flexibility		.15			
			-femininity		.12			
			CPI-dominance	Aptitude test composite score	.08			
			-cap. for status		.15			
			-sociability		.10			
			-social pressure		.04			
			-self-acceptance		.04			
			-well-being		.02			
			-responsibility		.14			
			-socialization		.09			
			-self-control		.00			
			-tolerance		.14			
			-good impression		.00			
			-communality		−.02			
			-ach. by conformity		.15			
			-ach. by independence		−.33			
			-intellectual efficiency		.26			
			-psychological minded		.04			
			-flexibility		.15			
			-femininity		.03			

c. The median r of three different samples.

Predictive validity of personality measures (continued)

Study and setting	Class or year	N	Personality measure	Outcome predicted	Predictive correlations r[c]	R_1	R_2	ΔR
Watson (1967) (continued)	1957–1960	100	CPI-dominance	GPA with aptitude composite partialed out	.15			
			-cap. for status		.04			
			-sociability		.16			
			-social presence		.07			
			-self-acceptance		.12			
			-well-being		.07			
			-responsibility		.16			
			-socialization		−.01			
			-self-control		.00			
			-tolerance		.07			
			-good impression		.00			
			-communality		.02			
			-ach. by conformity		.16			
			-ach. by independence		.23			
			-intellectual efficacy		.20			
			-psychological minded		.24			
			-flexibility		.04			
			-femininity		.10			

c. The median r of three different samples.

Other titles in this series:

The Validation of a Biographical Inventory as a Predictor of College Success. Research Monograph No. 1. Anne Anastasi, Martin J. Meade, and Alexander A. Schneiders. (200080) 1960, 86 pages, $5.

Organizational Measurement and Its Bearing on the Study of College Environment. Research Monograph No. 2. Allen H. Barton; introduction by Paul F. Lazarsfeld. (254730) 1961, 91 pages, $5.

Career Development: Choice and Adjustment. Research Monograph No. 3. David V. Tiedeman and Robert P. O'Hara. (213810) 1963, 115 pages, $5.

Career Development: Self-Concept Theory. Research Monograph No. 4. Donald E. Super, Reuben Starishevsky, Norman Matlin, and Jean Pierre Jordaan. (213815) 1963, 104 pages, $6.

Personality Measures in Admissions: Antecedent and Personality Factors as Predictors of College Success. Research Monograph No. 5. Morris I. Stein. (270500) 1963, 76 pages, $5.

The Measurement of Writing Ability. Research Monograph No. 6. Fred I. Godshalk, Frances Swineford, and William E. Coffman. (251700) 1966, 92 pages, $6.50.

Using Self-Reports to Predict Student Performance. Research Monograph No. 7. Leonard L. Baird. (251701) 1976, 92 pages, $5.

Population Validity and College Entrance Measures. Research Monograph No. 8. Hunter M. Breland. (270501) 1979, 104 pages, $7.50.

Orders accompanied by either a check made payable to the College Board or an institutional purchase order may be sent to:

College Board Publication Orders
Box 2815, Princeton, New Jersey 08541

ISBN: 0-87447-13
001